GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:10:50 ; Search time 3842.15 Seconds
                                              (without alignments)
                                              1829.132 Million cell updates/sec

Title: US-09-394-745-7565
Perfect score: 426
Sequence: 1 gggccgacccacgcgtccag..........catcgacacggtgcgagcct 426

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_om:*
                 20: em_or:*
                 21: em_ov:*
                 22: em_pat:*
                 23: em_ph:*
                 24: em_pl:*
                 25: em_ro:*
                 26: em_sts:*
                 27: em_sy:*

```
28:   em_un:*
29:   em_vi:*
30:   em_htgo_hum:*
31:   em_htgo_inv:*
32:   em_htgo_rod:*
33:   em_htg_hum:*
34:   em_htg_inv:*
35:   em_htg_rod:*
36:   em_htg_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	217	50.9	137462	8	AP002538	AP002538 Oryza sat
c	2	217	50.9	143515	8	AP002526	AP002526 Oryza sat
	3	98.4	23.1	114498	8	F3O9	AC006341 Arabidops
	4	93.6	22.0	80374	8	T8K14	AC007202 Arabidops
c	5	89.4	21.0	33270	3	CELR02F11	AF016439 Caenorhab
	6	82	19.2	138108	8	AP003231	AP003231 Oryza sat
c	7	81.8	19.2	86950	8	AC004218	AC004218 Arabidops
c	8	55.2	13.0	109016	8	ATT10K17	AL132977 Arabidops
	9	45	10.6	113193	1	AF357202	AF357202 Streptomy
	10	44.6	10.5	12829	1	AE004449	AE004449 Pseudomon
	11	44	10.3	1766	10	AF015304	AF015304 Rattus no
	12	44	10.3	35028	3	CELF56C9	U00063 Caenorhabdi
c	13	39	9.2	11548	1	AE005086	AE005086 Halobacte
	14	38.8	9.1	14713	1	RSCHECTOR	X80205 Rhodobacter
	15	38.6	9.1	3314	3	AY047566	AY047566 Drosophil
c	16	38.6	9.1	69061	2	AC012986	AC012986 Drosophil
c	17	38.6	9.1	168469	3	AC007886	AC007886 Drosophil
c	18	38.6	9.1	228448	3	AE003772	AE003772 Drosophil
	19	38	8.9	8991	1	SVI17268	Y17268 Streptomyce
	20	38	8.9	124182	2	AC091087	AC091087 Oryza sat
	21	38	8.9	144916	2	AP003505	AP003505 Oryza sat
c	22	38	8.9	155574	2	AC091090	AC091090 Oryza sat
c	23	38	8.9	160284	2	AP003437	AP003437 Oryza sat
	24	37.6	8.8	1929	6	A85321	A85321 Sequence 1
	25	37.6	8.8	1929	8	AF029858	AF029858 Sorghum b
	26	37.4	8.8	1591	10	AF305501	AF305501 Mus muscu
	27	37	8.7	10565	1	AE004621	AE004621 Pseudomon
	28	37	8.7	229896	14	AF232689	AF232689 Rat cytom
c	29	36.8	8.6	2982	1	AF134837	AF134837 Amycolato
	30	36.8	8.6	198677	1	AE001863	AE001863 Deinococc
	31	36.6	8.6	1377	9	HSU53143	U53143 Human inwar
	32	36.6	8.6	1788	9	HUMHCIR	L36069 Human high
	33	36.6	8.6	10029	1	AE008083	AE008083 Agrobacte
c	34	36.6	8.6	194780	2	AC068418	AC068418 Homo sapi
	35	36.2	8.5	33517	1	SC10B7	AL355752 Streptomy
	36	36.2	8.5	80609	1	AF116907	AF116907 Rhodococc
	37	36.2	8.5	80610	1	AP001204	AP001204 Rhodococc
	38	36	8.5	1998	1	STMHRDD	M90413 Streptomyce

```
     39     36    8.5    2682    6   E04309              E04309 DNA encodin
  c  40     36    8.5  349116    1   AP003003            AP003003 Mesorhizo
     41   35.8    8.4    1880   10   AF257189            AF257189 Mouse/rat
     42   35.8    8.4    1886   10   AF257188            AF257188 Mouse/rat
     43   35.8    8.4    1979   10   BC006812            BC006812 Mus muscu
     44   35.8    8.4    2013   10   AF131212            AF131212 Mus muscu
     45   35.8    8.4    2071   10   BC004828            BC004828 Mus muscu
```

                              ALIGNMENTS


```
RESULT    1
AP002538/c
LOCUS       AP002538   137462 bp    DNA            PLN       27-JUL-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0408F06.
ACCESSION   AP002538
VERSION     AP002538.2  GI:9558455
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0408F06.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 137462)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0408F06
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 137462)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jul 28, 2000 this sequence version replaced gi:8698576.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
            GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the coding regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RGP clone ID.
            This sequence of P0408F06 clone has an overlap with P0504H10 clone,
            DDBJ:AP002526 at the 3' end. The sequence of this clone ends at the
            position 42,574 of P0504H10. Detailed information on assemble
            quality together with annotation of this entry at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .137462
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
```

```
                    /db_xref="taxon:4530"
                    /chromosome="1"
                    /clone="P0408F06"
      CDS           join(2397. .2537,2639. .2810,4254. .4355,4465. .4604,
                    4709. .4801,4879. .4946,5037. .5110,5218. .5354)
                    /note="ESTs D48949(S15541),AU097625(S15541) correspond to
                    a region of the predicted gene.
                    Similar to Arabidopsis thaliana chromosome I BAC T22E19;
                    putative bifunctional nuclease (AC016447)"
                    /codon_start=1
                    /protein_id="BAB03377.1"
                    /db_xref="GI:9558456"
                    /translation="MALAAPLLRLRLLPLAAFVSVVSLTAAPRRAEAWGKQGHIIVCK
                    IAEKYLSEKAAAAVEELLPESAGGELSTVCPWADEVRFHYYWSRPLHYANTPQVCNFK
                    YSRDCHNSRHQQGMCVVGAINNYTDQLYSYGDSKSSYNLTESLMFLAHFVGDVHQPLH
                    VGFEEDEGGNTIKVHWYRRKENLHHVWDNSIIETAMKDFYNRSLDTMVEALKMNLTDG
                    WSEDISHWENCGNKKETCANDYAIESIHLSCNYAYKDVEQDITLGDDYFYSRYPIVEK
                    RLAQAGIRLALILNRIFGEDKPDGNVIPLQVQ"
      CDS           complement(join(5944. .5958,6006. .6111,6225. .6313,
                    6393. .6521,6608. .6708,6808. .7000,7084. .7119))
                    /note="EST D23006(C1998) corresponds to a region of the
                    predicted gene.
                    Similar to Synechocystis sp. PCC6803 complete genome;
                    hypothetical protein (D90915)"
                    /codon_start=1
                    /protein_id="BAB03378.1"
                    /db_xref="GI:9558457"
                    /translation="MAALLLLSSAARVGVAAPLALRQQRPVVLPGGQLRTGSGAGAAS
                    AWAARPLRPELAAVSRPAVPARGRAPLFRPRAWMASSQIASSAFTWGTIAVLPFYTLM
                    VVAPNADVTKRAVDSSAPYVALGILYAYLLYLSWTPDTLRAMFASKYWLPELTGIVRM
                    FASEMTVASAWIHLLAVDLFAARQVYHDGIKNNIETRHSVSLCLLFCPIGIATHVLTK
                    VHIA"
      CDS           complement(join(8173. .8409,8756. .8887,8968. .9047,
                    9150. .9279,9367. .9499,9607. .9728,9843. .9980,10724.
      .10909,
                    11807. .11944,12299. .12763))
                    /note="ESTs D22655(C0749),AU097597(C12421),C26485(C12421),
                    AU0976(C0749),C24828(S15393) correspond to a region of the
                    predicted gene.
                    Similar to Arabidopsis thaliana chromosome I BAC F15H11;
                    unknown protein (AC008148)"
                    /codon_start=1
                    /protein_id="BAB03379.1"
                    /db_xref="GI:9558458"
                    /translation="MAMDDLAGSSSSSSAMDAVVADPSHGWQKVTYPKRHRKQGAAAL
                    PSAAAPDLGFLPNGGGKVNVFEAVDRNAEKRHRALLAARDAADPDAARIAAATASAYS
                    DDDDDSDEAQATRPEGEVKKPKVKKPKKPKVTVAEAAALIDAENLAAHLVQISESYEN
                    QQDIQLMRFADYFGRSFASVSAAQFPWAKMFKESLVSKMVDIPLCHIPEPVRNTASDW
                    INQRSPDALGDFVMWCIDSIMSELSGQAVGAKGSKKAAQQTPRAQVAIFVVLALTVRR
                    KPEVLTNVLPKIMGNNKYLGQEKLPIIVWVIAQASQGDLVTGMFCWAHFLFPTLCAKP
                    SGNPQTRDLVLQLLERILSAPKARGILLNGAVRKGERLIPPVTFDLFMRAAFPVSSAR
                    VKATERFEAAYPTIKELALAGPPGSKTVKQAAQQLLPLCVKAMQENNADLTGESAGVF
                    IWCLTQNAESYKLWERLHPENVEASVVVLSTIVTKWSELSHKLSAESLKVTLKNLRTK
                    NEAALEAATDSGKQASIKAADKYSKEILGRLSRGGACLKGSLLVITLAVAAGFVLSPN
                    LEIPSDWDKLQAMVASHLSF"
      CDS           complement(join(18210. .18406,18420. .18576))
                    /note="hypothetical protein"
```

```
                          /codon_start=1
                          /protein_id="BAB03380.1"
                          /db_xref="GI:9558459"
                          /translation="MATGDATATGDATATAERRRDGDAAMGRRGAARGLARGDGGTTA
                          TGDAIVTGRVGLEAMRPETATARGTTRREGARQWRARQRRSDGARRLEARGARGRGDG
                          TGARRGRGAGHEATT"
        LTR               21126. .21841
                          /note="5' LTR"
        CDS               join(22727. .23791,24110. .24570,24649. .25022,25206.
.25340,

                          25516. .26239,26420. .28186)
                          /note="Similar to Oryza sativa chromosome 1 PAC P0003H10;
                          Gypsy-Ty3 type retrotransposon RetroSor1(AP000815)
                          probably inactive because one bp frameshift insertions and
                          stop codon are included in CDS and initiation codon was
                          not found."
                          /pseudo
                          /codon_start=1
        CDS               complement(join(28806. .29330,29350. .29361))
                          /note="Similar to Oryza sativa chromosome 5 PAC P0699E04;
                          unknown protein (AP001111)"
                          /codon_start=1
                          /protein_id="BAB03381.1"
                          /db_xref="GI:9558460"
                          /translation="MMSQGNFYIVGRRRRRRIHSPRCRRRRCNDCRRDRRRRARVLQT
                          MGNGEWGTEMVEPDQGGLAQRLSEMTGALERLPEELEETIKSSSRDLARGAVELVLAS
                          YQARDPDFSPWAALEEFPPGTEDGARAKVRDATDHIVHSFEGTAPRLAFALDFDEEGS
                          DDGADDSDDEADVPGASE"
        CDS               join(30218. .30424,30525. .32092)
                          /note="Similar to Zea mays mudrA protein (M76978)
                          probably inactive because one bp frameshift deletions and
                          stop codons are included in CDS."
                          /pseudo
                          /codon_start=1
        CDS               complement(join(33003. .33173,33376. .33447,34184. .34324,
                          34465. .34491))
                          /note="hypothetical protein"
                          /codon_start=1
                          /protein_id="BAB03382.1"
                          /db_xref="GI:9558461"
                          /translation="MEIDSRRDTSCYEFHELLPRCLLSDPIRRNRIRRNHQPSPPATV
                          LPRLLQEGKLAETDQRGNEFGGELYQWQWSLCNFEKNCRCALWIWEDLNEYVEEMVAY
                          CHADEYDYLRETCDSLRQFIADQRHAYLSVVSLG"
        CDS               complement(join(42733. .42925,42984. .43069))
                          /note="hypothetical protein"
                          /codon_start=1
                          /protein_id="BAB03383.1"
                          /db_xref="GI:9558462"
                          /translation="MAELVPIISTNSSSIALSSSYIYGGSMPHGWMVQLNDANQTPVA
                          IASKCKKQLSTMKKGSHLLSPEEEKEEDEDGIDRIHTKIGSLIEIGIM"
        LTR               44995. .46614
                          /note="5' LTR"
        CDS               46741. .50973
                          /note="Similar to Zea mays retrotransposon Opie-2
                          (T04112)"
                          /codon_start=1
                          /protein_id="BAB03384.1"
```

```
                         /db_xref="GI:9558463"
                         /translation="MEAYLQSQGHNVWNKVKSPYTVPDDADITPANMAQVDFNYRARN
                         AIIGGISSGEFNRVQHHKSAHDMWTALCNFHEGNNDIQLVRQNQFHKEYQRFEMHPGE
                         SIDSYFKRFGEIVSKLRSVGKEFSDNDNARHLLNCLDYGVWEMKVTSITESAPLSDLT
                         MDKLYSKLKTHEMDVFHRKGLKHSMALVADPSGSTSSNDSAFVCGGFSLAALHSVTEE
                         QLEKIPEDDLALFARKFSRAYKNVRNKKRGKTNEPFVCFECGEPNHIRVNCPKLKKKS
                         DKTTKKPEGQGRKGKNDLMKKAIHKVLAALEEVQLSDIDSDDDDQEKGDKDFSGMCCL
                         ANNEDFINLCLMALEDKDDSSEHPEDFGVGRSNSWLVDSGCSRHMTGEAKWFTSLTRA
                         SGDETITFGDASSGRVMAKGTIKVNDKFMLKDVALVSKLKYNLLSVSQLCDENLEVRF
                         KKDRSRVLDASESPVFDISRVGRVFFANFDSSAPGPSRCLVASENRDLFFWHRRLGHI
                         GFDHLSRISGMDLIRGLPKLKAPKDLVCAPCRHGKMTSSSHKPVTMVMTDGPGQLLHM
                         NTVGPARVQSVGGKWYVLVVVDDFSRYSWVYFLESKEETFGFFQSLARSLALEFPGAL
                         RAIRSDNGSEFKNSAFESFCDSSGVEHQFSSPYVPQQNGVVERKNRTLVEMARTMLDE
                         FTTPRKFWTEAISAACFISNRVFLRTILHKTPYELRFGRRPKVSHLRVFGCKCFVLKS
                         GNLDKFESRSLDGIFLGYATHSRAYRVYVLSTNKIVETCEVTFDEASPGARPEISGVL
                         DESIFVDEDSDDDDDDSIPPPLDSTPPVQETGSPSTTSPSGDAPTTSSSAAEEIDGGT
                         SGPTAPRHIQNRHPPDSMIGGLGERVTRNRSYDLVNSAFVASFEPKNVCHALSDENWV
                         NAMHEELENFERNKVWSLVEPPLGFNVIGTKWVFKNKLGEDGSIVRNKARLVAQGFTQ
                         VEGLDFEETFAPVARLEAIRILLAFAASKGFKLFQMDVKSAFLNGVIEEEVYVKQPPG
                         FENPKFPNHVFKLDKALYGLKQAPRAWYERLKTFLLQNGFEMGAVDKTLFTLHSGIDF
                         LLVQIYVDDIIFGGSSHALVAQFSDVMSREFEMSMMGELTFFLGLQIKQTKEGIFVHQ
                         TKYSKELLKKFDMADCKPIATPMATTSSLGPDEDGEEVDQREYRSMIGSLLYLTASRP
                         DIHFSVCLCARFQASPRTSHRQAVKRIFRYIKSTLEYGIWYSCSSALSVRAFSDADFA
                         GCKIDRKSTSGTCHFLGTSLVSWSSRKQSSVAQSTAEAEYVAAASACSQVLWMISTLK
                         DYGLSFSGVPLLCDNTSAINIAKNPVQHSRTKHIEIRYHFLRDNVEKGTIVLEFVESE
                         KQLADIFTKPLDRSRFEFLRSELGVIHPYGLI"
         LTR             52285. .53916
                         /note="3' LTR"
         CDS             join(61390. .61510,61761. .61886,62132. .62213,62343.
 .62805)

                         /note="EST C26936(C50482) corresponds to a region of the
                         predicted gene.
                         Similar to Arabidopsis thaliana zinc finger protein
                         (L39649)"
                         /codon_start=1
                         /protein_id="BAB03385.1"
                         /db_xref="GI:9558464"
                         /translation="MNSSRRQEGSPLDLNNLPDEFGKQTVESSTTTAASSAEASRVTK
                         KKSNGGKDEAGKVYECRFCSLKFCKSQALGGHMNRHRQERETETLNRARQLVFGNDSL
                         AAVGAQLNFRDVNMGGGGAAAPPPTMQMGGGGFRGGGVGGDPCIPLRPVQPRLSPPQP
                         PPYHHYLYTTTAPPSALHPMSYPATYPAPPRHQQPAAVGDYVIGHAVSAGDALVAPPP
                         PPHRASFSCFGAPLAAPPANVQPDNGNCNCSFGCGHSNRNVNAAS"
         CDS             complement(join(69176. .69288,71061. .71208))
                         /note="hypothetical protein"
                         /codon_start=1
                         /protein_id="BAB03386.1"
```

```
    Query Match             50.9%;   Score 217;  DB 8;  Length 137462;
    Best Local Similarity   71.6%;   Pred. No. 2.9e-43;
    Matches   336;  Conservative    0;  Mismatches   60;  Indels   73;  Gaps     1;

Qy     31 ccaatcaggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgt 90
           ||   |||||||||   |   ||| |   || | | ||||| |||||| ||    |||||||
Db 110724 CCTTACAGGAGCACCAAGTGTTCGACCACAGGAAGGAGCTGTGGATGATCGGCAGCATGT 110665

Qy     91 cctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgaga 150
           ||||  || || ||  |||||||||||||||| |||||||||||||||  ||||||||| ||||
```

```
Db 110664 CCTCAGTCGCAGTGGTGAAGTTCTTCCTGATGCTCTACTGCCGGTCGTTCAAGAACGAGA 110605

Qy    151 tcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctgg 210
          ||||||| ||||||||| |||||||||||||||||||| |||||| |||||| ||||||||| |
Db 110604 TCGTGAGAGCCTACGCGCAGGACCATTTCTTCGACGTGATCACCAACTCGGTCGGCCTCG 110545

Qy    211 tctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatact-- 268
          ||    ||||| |||||| |||||||||||||||||||||| || || || |||||||||
Db 110544 TCAGCGCGCTCCTCGCCGTCCGGTACAAATGGTGGATGGATCCGGTCGGAGCCATACTGG 110485

Qy    269 ------------------------------------------------------------ 268

Db 110484 TGAGTGCCCCCATTGCTGCCTGCCTGCCACTCTGCTAGCTACTCCATGTGAGAATTAATG 110425

Qy    269 -----------gatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaac 317
                     ||||||| |||||||||||||||||||||| || |||||| |||||||||
Db 110424 GTGGATATGCAGATCGCGGTGTACACGATCACGACGTGGGCTCGGACGGTGGTGGAGAAC 110365

Qy    318 gtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttg 377
          || || || ||||| |||| |||||||||||| |||||||||||||||| ||||||| ||
Db 110364 GTGGGGACGCTGATCGGCAGGTCGGCGCCGGCGGAGTACCTGACGAAGCTGACGTACCTG 110305

Qy    378 atctggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
          || |||||||||||| |||||||||| |||||||||||||||| | ||||
Db 110304 ATATGGAACCACCACGAGGAGATCCGGCACATCGACACGGTGAGGGCCT 110256


RESULT    2
AP002526/c
LOCUS       AP002526   143515 bp    DNA              PLN       11-JUL-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0504H10.
ACCESSION   AP002526
VERSION     AP002526.1  GI:8570080
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0504H10.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 143515)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0504H10
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 143515)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PIR,SWISSPROT,
```

GENPEPT, PDB) from MAFF DNAbank and the cDNA sequence database at
RGP. Protein similarities of the coding regions were searched
against NRP with BLASTP2.0. ESTs represent the identified cDNA
sequences using BLASTN2.0 with the corresponding DDBJ accession no.
and RGP clone ID. Detailed information on assemble quality together
with annotation of this entry at
http://rgp.dna.affrc.go.jp/genomicdata/GenomeFinished.html.

```
FEATURES             Location/Qualifiers
     source          1. .143515
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0504H10"
     CDS             join(3250. .3298,6051. .6550)
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA99360.1"
                     /db_xref="GI:9049405"
                     /translation="MASNNVLFIMLARESVPEGRGGREGGRERRRLRRRRHGMREAAA
                     AALRGRREGAAVAAEGRGGRRSMPAEGRRRLRRHRRGMREAVVEPPAQARGRREEAAV
                     AAPDPLTAARAPPPGLTAATPRHRRHQGRRRLAFCRLSLWWIRGGEDKNTLSPLVVVD
                     VITKAGVGSPSRLWPPLPVVNP"
     CDS             complement(join(11974. .12275,13114. .13225))
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA99361.1"
                     /db_xref="GI:9049406"
                     /translation="MDGLRWRPAVAHVSYPSSSSSSSSLGPGKWTPGEAGSPSSMPPP
                     PATTPLPRRRLALILCLAWALWLHGGGGGISLADAFQAPTPARLSSGSSYAVGSRPVP
                     AAAPRWSSSSASEAAARFADDKRRIPSCPDALHNR"
     CDS             complement(join(15180. .15524,15598. .15828,15925. .16005,
                     16197. .16442,16590. .16705,18078. .18234))
                     /note="ESTs AU032649(S13048),D47505(S13048) correspond to
                     a region of the predicted gene.
                     Similar to Arabidopsis thaliana chromosome II BAC F12L6;
                     unknown protein. (AC004218)"
                     /codon_start=1
                     /protein_id="BAA99362.1"
                     /db_xref="GI:9049407"
                     /translation="MGSRGRRGGGERETETEEDETWKLRVGDDFTVPERFHRKPPFFS
                     RIFPAGSHGKHRKIAKYYKKQENLLKDFSEMETMNEIGSLDQNAPTEEELRQMAKGER
                     LAINLSNIINLILFIGKVLASVESLSMAVIASTLDSLLDLLSGFILWFTAHAMKKPNK
                     YSYPIGKRRMQPVGIIVFASVMGTLGFQVLIESGRQLITNEHQVFDHRKELWMIGSMS
                     SVAVVKFFLMLYCRSFKNEIVRAYAQDHFFDVITNSVGLVSALLAVRYKWWMDPVGAI
                     LIAVYTITTWARTVVENVGTLIGRSAPAEYLTKLTYLIWNHHEEIRHIDTVRAYTFGT
                     HYFVEVDIVLPGDMPLSHAHDIGESLQEKLEQLPEVERAFVHVDFEFTHRPEHKAEV"
     CDS             complement(join(23230. .23438,24747. .24756))
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAA99363.1"
                     /db_xref="GI:9049408"
                     /translation="MTTVIQNLSPLSTHALHGRPVEMTTGEIEVSIAALATKKALQEA
                     FDVLTAACSPFTWGDLNSYISSLQSSID"
     CDS             complement(join(26770. .27247,27516. .27541,28770. .28775))
                     /note="hypothetical protein"
                     /codon_start=1
```

```
                            /protein_id="BAA99364.1"
                            /db_xref="GI:9049409"
                            /translation="MEHMFRYQNVLKHSSFLFYYTTNTATRTHPPPNVYAVTNTTTMD
                            ATTTAKRKRPAASDIADDAPTTVDEVSDAEVEEFYAILRRMRDATRRLGARPPPPRAP
                            AWRPSFSWEDFADAPPKQAPPPPQQPADHERVAENATPPRRPAPGLDLNVEPPSDAPA
                            TPRSARAPA"
            CDS             join(29697. .29733,30017. .30158,31506. .31544,32422.
  .32559,

                            36042. .36172,36223. .36326,37134. .37259)
                            /note="hypothetical protein"
                            /codon_start=1
                            /protein_id="BAA99365.1"
                            /db_xref="GI:9049410"
                            /translation="MSFGNHIINWIKAHYEFPFHRDKTAQLYHPRRYYVPRFQDKEAV
                            KLIFTERFPNATRLCLSTGSQQMLARESSSSYKISMYASMEKSLPASTVKHFSKRKTK
                            REDNAGVELYRSGLRRDGRIQPSGPARRQPRFRRRGPLQMGCPLLSFYMLLGRIGWAN
                            CRDGTVRRERYSGQQHGHVQRASNPKPSRERWSRTPHDLQLMMMVDELMLSEQKGWRT
                            RPGTRLAAVVRQSGKILWQR"
            CDS             join(39790. .40034,40493. .41151,42097. .43982)
                            /note="EST AU092739(C53221) corresponds to a region of the
                            predicted gene.
                            Similar to Arabidopsis thaliana alpha-xylosidase precursor
                            (AF087483)"
                            /codon_start=1
                            /protein_id="BAA99366.1"
                            /db_xref="GI:9049411"
                            /translation="MLASLSSSSRAAISCIPLCLLFLTLASSNGVFAAAPPKVGSGYK
                            LVSLVEHPEGGALVGYLQVKQRTSTYGPDIPLLRLYVKHETKDRIRVQITDADKPRWE
                            VPYNLLQREPAPPVTGGRITGVPFAAGEYPGEELVFTYGRDPFWFAVHRKSSREALFN
                            TSCGALVFKDQYIEASTSLPRDAALYGLGENTQPGGIRLRPNDPYTIYTTDISAINLN
                            TDLYGSHPVYVDLRSRGGHGVAHAVLLLNSNGMDVFYRGTSLTYKVIGGLLDFYLFSG
                            PTPLAVVDQYTSMIGRPAPMPYWAFGFHQCRWGYKNLSVVEGVVEGYRNAQIPLDVIW
                            NDDDHMDAAKDFTLDPVNYPRPKLLEFLDKIHAQGMKYIVLIDPGIAVNNTYGVYQRG
                            MQGDVFIKLDGKPYLAQVWPGPVYFPDFLNPNGVSWWIDEVRRFHDLVPVDGLWIDMN
                            EASNFCTGKCEIPTTHLCPLPNTTTPWVCCLDCKNLTNTRWDEPPYKINASGQTARLG
                            FNTIATSATHYNGILEYNAHSLYGFSQAIATHQALQGLQGKRPFILTRSTFVGSGAYA
                            AHWTGDNKGTWENLRYSISTMLNFGIFGMPMVGADICGFYPQPTEELCNRWIELGAFY
                            PFSRDHANFASPRQELYVWESVAKSARNALGMRYRLLPYLYTLNYQAHLTGAPVARPV
                            FFSFPDFTPCYGLSTQYLLGASVMVSPVLEQGATSVSAMFPPGSWYNLFDTTKVVVSR
                            GEGAVKLDAPLNEINVHVFQNTILPMQRGGTISKEARATPFTLVVAFPFGATEAEAEG
                            AVYVDDDERPEMVLAEGQATYVRFYATVRGKAVTVRSEVELGSYSLQKGLLIEKLSVL
                            GLEGTGRDLAVHVDGANATAIATSRPYFAGAEAELHGHRDVEGHKKSVMVEVGGLALP
                            LGKSFTMTWNMQIEA"
            CDS             complement(join(44513. .44628,44655. .44800,45604. .45695))
                            /note="hypothetical protein"
                            /codon_start=1
                            /protein_id="BAA99367.1"
                            /db_xref="GI:9049412"
                            /translation="MCRAEKYAHFSWQNLNREASQFSVNQRRTNLVVHNYYASRDVTV
                            VVATHYPTFVFGPQMLFTGLSKVSDGLISLWATCCGLQITMARIIFLMNQACSNHRFV
                            SYQTTEKNSEHKTVK"
            CDS             complement(join(46824. .47117,47123. .47218))
                            /note="hypothetical protein"
                            /codon_start=1
                            /protein_id="BAA99368.1"
                            /db_xref="GI:9049413"
                            /translation="MVKPGIQVDPGLRRLRRRCGGFRRAAMGSCVVSRVSASSGTKGA
```

```
                              VRSFADRDRALRLQCNWPDLTTARRPLCVLGSGWLAGSASRGVPFGNLCSARARRGCS
                              CMAADSSSGLGAAWRDLAPRGSRVGGT"
        CDS                   complement(join(50933. .50968,51690. .51904,52049. .52127))
                              /note="hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99369.1"
                              /db_xref="GI:9049414"
                              /translation="MSAWKATWTPPVSPPLFFFSLSSLRRARRRMAAAMDEDFARAVE
                              DDLKLSKRLVLPGGRPRPAPSLLTRPLLTAPMCPATSHTYTAASTRPSSSHSTGVRMP
                              RMHYQMF"
        CDS                   complement(53670. .54194)
                              /note="ESTs AU076137(E30609),U076136(E30609) correspond to
                              a region of the predicted gene.
                              hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99370.1"
                              /db_xref="GI:9049415"
                              /translation="MKRTRAQQPKLQEGQDGGGAAGNANPKPQRRAKQPRQPKAASAA
                              AKKAAAAAAARESSSSSVGAGAAVTSAASSSCSSGADMAPTVPDVCGGGGGGAGYEAG
                              AATTVEWDLDGGLSNGLSWWTFGVEEEKLLGWFPFVEEDFRCLGARGDAEMAFDDDIW
                              RIHQIYEIPNYAAK"
        CDS                   complement(join(54570. .54596,54672. .54814,55810. .55895,
                              57161. .57258,57791. .57884,58516. .58577))
                              /note="hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99371.1"
                              /db_xref="GI:9049416"
                              /translation="MVNLGKEMEFHRNGCIKRCRSVEVDVIKLPIQIFRDARVFRERA
                              LCGSFTERPSFAAYLEEKLDVCEYLREMEEGERERTKVYDNHLYRRSDAAFMRLWIVV
                              FFAELQLRAQPDRQAHALVDAQPPQVSGGFHVSRPSITGRCYAHISWKKKTAKATHPG
                              PYNLVRFIN"
        CDS                   join(60308. .60494,60508. .60705,60733. .61172,65600.
  .65674)

                              /note="hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99372.1"
                              /db_xref="GI:9049417"
                              /translation="MAKAGARGAGQWWMKAAGMASAEGAKAAATDFATGSVSAAGGSR
                              VRRGEEVVSSGDGGGWWRGDVEADGMWRLSCGLVGGEVGGVCRRRRPGAVHAVAEGRR
                              GGRWRCGVAASGVELAGDDSGCRWTAATVGDRGCGWRRHGGLGRLAEGVANGCIWPAQ
                              QCFEERSETGLALRGVADGSGGRYDARGVAGEDGGWLRARGTADGGRPDWRERRTRWR
                              RPAWRERRGRRWRRRPRCEEELPIGVARSTAHEGLPAGGAGTQWSHMSTEVERWWSLV
                              SYEISFFIRVQKLFCQYPVKHMM"
        CDS                   complement(join(67737. .68066,69231. .69373,70526. .70781))
                              /note="hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99373.1"
                              /db_xref="GI:9049418"
                              /translation="MPIPEPDRTSEPDSEPMRPIQSPRVALVATETSRCARPTPCGPH
                              RLRIRPEVGPRRVKFRASAAPDLEWMKMPVLPLLSLRCHVSGQINGATPSMGVSFAAV
                              RFSLRRARATTKHAPTTTTTRRQAQPAGVRRRLPPTAARHWARARPSRPPAAWEGKGR
                              EEGRGWEGKGRRWRGRRREIDLRGRRGGGDEEREIDLSGWEGRTGRRGPGMGGGMGGG
                              GGEGDRSEREEREIDLRGRRGRSI"
        CDS                   complement(join(72289. .72685,76625. .76668))
                              /note="hypothetical protein"
                              /codon_start=1
                              /protein_id="BAA99374.1"
```

/db_xref="GI:9049419"
/translation="MGAYTVEVAFPCAVRATRASLRHPRAPASPTDTHCSAPMPATRH
GVATRFDQIQPVAGQIRRLQWRIYRRLAGRARCCHSSPSALAASPAPKRMCRRRRRPR

```
Query Match            50.9%;  Score 217;  DB 8;  Length 143515;
Best Local Similarity  71.6%;  Pred. No. 2.8e-43;
Matches  336;  Conservative   0;  Mismatches  60;  Indels  73;  Gaps   1;


Qy      31 ccaatcaggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgt 90
              ||  |||||||||   |  ||| |   || | | ||||| |||||| ||    |||||||
Db   15836 CCTTACAGGAGCACCAAGTGTTCGACCACAGGAAGGAGCTGTGGATGATCGGCAGCATGT 15777


Qy      91 cctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgaga 150
              ||||  || || ||  |||||||||||||| |||||||||||| |||||||||| ||||
Db   15776 CCTCAGTCGCAGTGGTGAAGTTCTTCCTGATGCTCTACTGCCGGTCGTTCAAGAACGAGA 15717


Qy     151 tcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctgg 210
              ||||||| |||||||| |||||||||||||||||||| ||||| |||||| ||||||| |
Db   15716 TCGTGAGAGCCTACGCGCAGGACCATTTCTTCGACGTGATCACCAACTCGGTCGGCCTCG 15657


Qy     211 tctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatact-- 268
              ||     |||||  ||||| |||||||||||||||||||||||||| || || || |||||||||
Db   15656 TCAGCGCGCTCCTCGCCGTCCGGTACAAATGGTGGATGGATCCGGTCGGAGCCATACTGG 15597


Qy     269 ------------------------------------------------------------ 268

Db   15596 TGAGTGCCCCCATTGCTGCCTGCCTGCCACTCTGCTAGCTACTCCATGTGAGAATTAATG 15537


Qy     269 -----------gatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaac 317
                        |||||||  |||||||||||||||||||||| || |||||| |||||||||
Db   15536 GTGGATATGCAGATCGCGGTGTACACGATCACGACGTGGGCTCGGACGGTGGTGGAGAAC 15477


Qy     318 gtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttg 377
              || || || |||||| |||| ||||||||||| |||||||||||||||||||| |||||| ||
Db   15476 GTGGGGACGCTGATCGGCAGGTCGGCGCCGGCGGAGTACCTGACGAAGCTGACGTACCTG 15417


Qy     378 atctggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
              || ||||||||||| |||||||||| |||||||||||||||| | ||||
Db   15416 ATATGGAACCACCACGAGGAGATCCGGCACATCGACACGGTGAGGGCCT 15368
```


RESULT   3
F3O9
LOCUS   F3O9   114498 bp   DNA   PLN   02-JUN-1999
DEFINITION Arabidopsis thaliana chromosome 1 BAC F3O9 sequence, complete
            sequence.
ACCESSION   AC006341
VERSION   AC006341.2 GI:4887257
KEYWORDS   HTG.
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 114498)
  AUTHORS   Vysotskaia,V.S., Schwartz,J., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
            Lee,J., Li,J., Kremenetskaia,I., Liu,A., Luros,J., Gonzalez,A.,

```
            Altafi,H., Araujo,R., Chao,Q., Conn,L., Conway,A.B., Dunn,P.,
            Hansen,N., Huizar,L., Kim,C., Palm,C., Rowley,D., Shinn,P.,
            Walker,M., Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
  TITLE     The sequence of BAC F3O9 from Arabidopsis thaliana chromosome 1
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 114498)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   3  (bases 1 to 114498)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   4  (bases 1 to 114498)
  AUTHORS   Theologis.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Plant Gene Expression Center, 800 Buchanan
            St., Albany, CA 94710, USA
COMMENT     On May 25, 1999 this sequence version replaced gi:4139327.
            The sequence of BAC F3O9 from Arabidopsis thaliana chromosome 1.
FEATURES             Location/Qualifiers
     source          1. .114498
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F3O9"
     gene            246. .901
                     /gene="F3O9.1"
     CDS             join(246. .275,353. .462,608. .692,785. .901)
                     /gene="F3O9.1"
                     /note="Similar to gb|Y12014 RAD23 protein isoform II from
                     Daucus carota.  This gene is probably cut off.  EST
                     gb|AA651284 comes from this gene."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD34676.1"
                     /db_xref="GI:4966345"
                     /translation="MVNSNPQILQPMLQELGKQNPQLLRLIQENQAEFLQLLNEPYEG
                     SDGDVDIFDQPDQEMPHSVNVTPEEQESIERLEAMGFDRAIVIEAFLSCDRNEELAAN
                     YLLEHSADFED"
     gene            complement(1361. .2444)
                     /gene="F3O9.2"
     CDS             complement(join(1361. .1387,1506. .1625,1899. .2444))
                     /gene="F3O9.2"
                     /note="ESTs gb|T04357 and gb|AA595092 come from this
                     gene."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD34673.1"
                     /db_xref="GI:4966342"
                     /translation="MGLNSKAEVAKSRKNAAEAEQKDRQTREKEEQYWREAEGPKSKA
                     VKKREEEAEKKAETAAKKLEAKRLAEQEEKELEKALKKPDKKANRVTVPVPKVTEAEL
                     IRRREEDQVALAKKAEDSKKKQTRMAGEDEYEKMVLVTNTNRDDSLIEAHTVDEALAR
                     ITVSDNLPVDRHPEKRLKASFKAYEEVELPRLKSEKPGLTHTQYKDLIWKMWKKSPDN
```

```
                     PLNQAAAAAANE"
     gene            3662. .5562
                     /gene="F3O9.3"
     CDS             join(3662. .3916,4107. .4475,4552. .4669,4755. .4978,
                     5053. .5562)
                     /gene="F3O9.3"
                     /note="Is a member of PF|00481 Protein phosphatase 2C
                     family."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD34674.1"
                     /db_xref="GI:4966343"
                     /translation="MGLCHSKIDKTTRKETGATSTATTTVERQSSGRLRRPRDLYSGG
                     EISEIQQVVGRLVGNGSSEIACLYTQQGKKGTNQDAMLVWENFCSRSDTVLCGVFDGH
                     GPFGHMVSKRVRDMLPFTLSTQLKTTSGTEQSSSKNGLNSAPTCVDEEQWCELQLCEK
                     DEKLFPEMYLPLKRALLKTCQQMDKELKMHPTINCFCSGTTSVTVIKQGKDLVVGNIG
                     DSRAVLATRDQDNALVAVQLTIDLKPDLPSESARIHRCKGRVFALQDEPEVARVWLPN
                     SDSPGLAMARAFGDFCLKDYGLISVPDINYHRLTERDQYIILATDGVWDVLSNKEAVD
                     IVASAPSRDTAARAVVDTAVRAWRLKYPTSKNDDCAVVCLFLEDTSAGGTVEVSETVN
                     HSHEESTESVTITSSKDADKKEEASTETNETVPVWEIKEEKTPESCRIESKKTTLAEC
                     ISVKDDEEWSALEGLTRVNSLLSIPRFFSGELRSSSWRKWL"
     gene            complement(10192. .11349)
                     /gene="F3O9.4"
     CDS             complement(join(10192. .10323,10518. .10562,10642. .10742,
                     10834. .11039,11135. .11349))
                     /gene="F3O9.4"
                     /note="ESTs gb|F15498, gb|H37515, gb|T41906, gb|T22448,
                     gb|W43356 and gb|T20739 come from this gene."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD34675.1"
                     /db_xref="GI:4966344"
                     /translation="MASSSDSWMRAYNEALKLSEEINGMISERSSSAVTGPDAQRRAS
                     AIRRKITIFGNKLDSLQSLLAEIHGKPISEKEMNRRKDMVGNLRSKANQMANALNMSN
                     FANRDSLLGPDIKPDDSMSRVTGMDNQGIVGYQRQVMREQDEGLEQLEGTVMSTKHIA
                     LAVSEELDLQTRLIDDLDYHVDVTDSRLRRVQKSLAVMNKNMRSGCSCMSMLLSVLGI
                     VGLAVVIWMLVKYM"
     gene            11987. .13706
                     /gene="F3O9.5"
     CDS             join(11987. .12044,12084. .12224,12408. .12954,13133.
  .13706)
                     /gene="F3O9.5"
                     /note="Contains two PF|01344 Kelch motif domains."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD34677.1"
                     /db_xref="GI:4966346"
                     /translation="MPVSSVSSLPCQNPEFLSDFSVSIIVNGVFVSGKVINFEARFWC
                     SKASSRRGSEDTFTISRVRFGSCLPDDLALRCIAKLSHGYHGVLECVSRGWRDLVRGA
                     DYSCYKARNGWSGSWLFVLTERSKNQWVAYDPEADRWHPLPRTRAVQDGWHHSGFACV
                     CVSNCLLVIGGCYAPSVSSFPHQKPVVTKDVMRFDPFKKQWKMVASMRTPRTHFACTS
                     VSGKVYVAGGRNLTHSRGIPSAEVYDPVADRWEELPAMPRPQMDCSGLSYRGCFHVLS
                     DQVGFAEQNSSEVFNPRDMTWSTVEDVWPFSRAMQFAVQVMKNDRVYTIVDWGESLIK
                     TRDTDEGEWYNVGSVPSVVLPNHPRELEAFGYGFAALRNELYVIGGKVLKWEESGAGR
                     FDIVRLPVVRVCNPLDRPLNWRETKPMCIPAGGSIIGCVSLEESSPP"
     gene            complement(14717. .17027)
                     /gene="F3O9.6"
```

```
CDS                complement(join(14717. .15924,15993. .16097,16178. .17027))
                   /gene="F3O9.6"
                   /note="Similar to gb|AJ012423 wall-associated kinase 2
                   from Arabidopsis thaliana."
                   /codon_start=1
                   /evidence=not_experimental
                   /protein_id="AAD34678.1"
                   /db_xref="GI:4966347"
                   /translation="MGVDVKRFLVVMLLLRICEYAAASTFPLALRNCSDHCGNVSVPY
                   PFGIGKGCYKNKWFEIVCKSSSDQQPILLLPRIRRAVTSFNLGDPFSISVYNKFYIQS
                   PLKHSGCPNRDGYSSSSLNLKGSPFFISENNKFTAVGCNNKAFMNVTGLQIVGCETTC
                   GNEIRSYKGANTSCVGYKCCQMTIPPLLQLQVFDATVEKLEPNKQGCQVAFLTQFTLS
                   GSLFTPPELMEYSEYTTIELEWRLDLSYMTSKRVLCKGNTFFEDSYQCSCHNGYEGNP
                   YIPGGCQDIDECRDPHLNKCGKRKCVNVLGSYRCEKTWPAILSGTLSSGLLLLIFGMW
                   LLCKANRKRKVAKQKRKFFQRNGGLLLQQQTSFLHGSVNRTKVFSSNDLENATDRFNA
                   SRILGQGGQGTVYKGMLEDGMIVAVKKSKALKEENLEEFINEIILLSQINHRNVVKIL
                   GCCLETEVPILVYEFIPNRNLFDHLHNPSEDFPMSWEVRLCIACEVADALSYLHSAVS
                   IPIYHRDVKSTNILLDEKHRAKVSDFGISRSVAIDDTHLTTIVQGTIGYVDPEYLQSN
                   HFTGKSDVYSFGVLLIELLTGEKPVSLLRRQEVRMLGAYFLEAMRNDRLHEILDARIK
                   EECDREEVLAVAKLARRCLSLNSEHRPTMRDVFIELDRMQSKRKGTQSQAQNGEEHAH
                   IQIAMPESMSLSYSSPNIVVENSSFSLDTKPLMPHKTQ"
gene               18899. .23154
                   /gene="F3O9.7"
CDS                join(18899. .21385,21458. .21538,21719. .21871,21997.
.22160,

                   22271. .22418,22491. .22613,22684. .22771,22955. .23154)
                   /gene="F3O9.7"
                   /note="Contains PF|00069 Eukaryotic protein kinase domain.
                    ESTs gb|H37741, gb|T43005 and gb|AI100340 come from this
                   gene."
                   /codon_start=1
                   /evidence=not_experimental
                   /protein_id="AAD34679.1"
                   /db_xref="GI:4966348"
                   /translation="MDRNRPPHPFQQHAMEPGYVNDSVPQGFTPDQTGLSNANVRPNP
                   ADVKPGLHYSIQTGEEFSLEFLRDRVISQRSANPIAAGDINYPTGYNGHAGSEFGSDV
                   SRMSMVGNGIRQYERTNPPVHEFGNKLGHIHSAPEASLCQDRSLGNFHGYASSSASGS
                   LTAKVKVLCSFGGKILPRPGDSKLRYVGGETHIISIRKDISWQELRQKVLEIYYRTHV
                   VKYQLPGEDLDALVSVSCDEDLLNMMEEYNEMENRGGSQKLRMFLFSVSDLDGALLGV
                   NKSDVDSEFQYVVAVNDMDLGSRSNSTLNGLDSSSANNLAELDVRNTEGINGVGPSQL
                   TGIDFQQSSMQYSESAPPTSFAQYPQSIPHNGAFQFQQAVPPNATLQYAPSNPPSSSV
                   HYPQSILPNSTLQYPQSISSSSYGLYPQYYGETEQFPMQYHDHNSSNYSIPIPFPGQP
                   YPHPGITQQNAPVQVEEPNIKPETKVRDYVEPENRHILATNHQNPPQADDTEVKNREP
                   SVATTVPSQDAAHMLPPRRDTRQNTPVKPSTYRDAVITEQVPVSGEDDQLSTSSGTCG
                   LVHTDSESNLIDLDYPEPLQPTRRVYRSERIPREQLEMLNRLSKSDDSLGSQFLMSHP
                   QASTGQQEPAKEAAGISHEDSHIVNDVENISGNVVASNETLDKRTVSGGGIETEARNL
                   SHVDTERSHDIPEKQTSSGVLIDINDRFPQDFLSEIFAKALSDDMPSGANPYQHDGAG
                   VSLNVENHDPKNWSYFRNLADEQFSDRDVAYIDRTPGFPSDMEDGGEIARLHQVAPLT
                   ENRVDPQMKVTESEEFDAMVENLRTSDCEQEDEKSETRNAGLPPVGPSLADYDTSGLQ
                   IIMNDDLEELKELGSGTFGTVYHGKWRGSDVAIKRIKKSCFAGRSSEQERLTGEFWGE
                   AEILSKLHHPNVVAFYGVVKDGPGATLATVTEYMVDGSLRHVLVRKDRHLDRRKRLII
                   AMDAAFGMEYLHAKNIVHFDLKCDNLLVNLKDPSRPICKVGDFGLSKIKRNTLVSGGV
                   RGTLPWMAPELLNGSSSKVSEKVDVFSFGIVLWEILTGEEPYANMHYGAIIGGIVNNT
                   LRPTIPSYCDSDWRILMEECWAPNPTARPSFTEIAGRLRVMSTAATSNQSKPPAHKAS
                   K"
gene               complement(23491. .25496)
                   /gene="F3O9.8"
```

```
CDS             complement(join(23491. .23643,23824. .23922,24023. .24104,
                24189. .24392,24473. .24795,24882. .25496))
                /gene="F3O9.8"
                /note="Similar to gb|L13612 DEAD-box protein (dbp45A) from
                Drosophila melanogaster and is a member of PF|00270
                DEAD/DEAH box helicase family."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAD34681.1"
                /db_xref="GI:4966350"
                /translation="MEEPTPEEEGGITIMSKSRKNPKTVVNIQSQKLDSDQNTPQFEK
                FTNPNPSSDTTSATNFEGLGLAEWAVETCKELGMRKPTPVQTHCVPKILAGRDVLGLA
                QTGSGKTAAFALPILHRLAEDPYGVFALVVTPTRELAFQLAEQFKALGSCLNLRCSVI
                VGGMDMLTQTMSLVSRPHIVITTPGRIKVLLENNPDVPPVFSRTKFLVLDEADRVLDV
                GFQDELRTIFQCLPKSRQTLLFSATMTSNLQALLEHSSNKAYFYEAYEGLKTVDTLTQ
                QFIFEDKDAKELYLVHILSQMEDKGIRSAMIFVSTCRTCQRLSLMLDELEVENIAMHS
                LNSQSMRLSALSKFKSGKVPILLATDVASRGLDIPTVDLVINYDIPRDPRDYVHRVGR
```

Query Match   23.1%;   Score 98.4;   DB 8;   Length 114498;
Best Local Similarity 65.5%;   Pred. No. 3.2e-14;
Matches 144;   Conservative   0;   Mismatches   76;   Indels   0;   Gaps   0;

```
Qy      49 atttcaagttcaagcaagagctctggatggtcattagcatgtcctctgttgcggtcgtga 108
           || | ||   ||    | |||    |||||| |   || ||||  ||||||  |  ||||||
Db   34944 ATATGAATAGCACAGAGGAGAAATGGATGATTGGTATCATGGTCTCTGTCACCATCGTGA 35003

Qy     109 agttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgccc 168
           ||||  | || |||||| || |||  ||    ||  ||| || |||||| |||||||| ||  |
Db   35004 AGTTTCTTCTAATGCTTTATTGCAGAGGATTTCAGAACGAAATCGTTAGGGCCTATGCTC 35063

Qy     169 aggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctg 228
           | || ||  |||| |||||   |||| || ||   |||||| |  | || | ||  || |
Db   35064 AAGATCACCTCTTTGACGTCGTCACTAATTCAATCGGCTTAGCAACAGCTGTCTTGGCCG 35123

Qy     229 tccggtacaaatggtggatggaccctgttggcgccatact 268
           ||    | | | |||||||||| || |||    || || |||||
Db   35124 TCAAATTCTACTGGTGGATTGATCCTACCGGTGCTATACT 35163
```


RESULT   4
T8K14
LOCUS   T8K14   80374 bp   DNA   PLN   05-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC T8K14 sequence, complete
            sequence.
ACCESSION   AC007202
VERSION   AC007202.3 GI:12039264
KEYWORDS   HTG.
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 80374)
  AUTHORS   Vysotskaia,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
            Li,J., Kremenetskaia,I., Luros,J., Lee,J.M., Gonzalez,A.,
            Altafi,H., Araujo,R., Chao,Q., Conn,L., Conway,A.B., Dunn,P.,
            Hansen,N., Huizar,L., Kim,C., Palm,C., Rowley,D., Shinn,P.,

```
                  Walker,M., Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
      TITLE       Arabidopsis thaliana chromosome 1 BAC T8K14 sequence
      JOURNAL     Unpublished
  REFERENCE       2  (bases 1 to 80374)
      AUTHORS     Theologis,A.
      TITLE       Direct Submission
      JOURNAL     Submitted (03-APR-1999) Plant Gene Expression Center, 800 Buchanan
                  Street, Albany, CA 94710, USA
  REFERENCE       3  (bases 1 to 80374)
      AUTHORS     Theologis,A.
      TITLE       Direct Submission
      JOURNAL     Submitted (29-APR-1999) Plant Gene Expression Center, 800 Buchanan
                  Street, Albany, CA 94710, USA
  REFERENCE       4  (bases 1 to 80374)
      AUTHORS     Theologis.
      TITLE       Direct Submission
      JOURNAL     Submitted (15-MAY-1999) Plant Gene Expression Center, 800 Buchanan
                  St., Albany, CA 94710, USA
  REFERENCE       5  (bases 1 to 80374)
      AUTHORS     Theologis,A.
      TITLE       Direct Submission
      JOURNAL     Submitted (05-JAN-2001) Plant Gene Expression Center, 800 Buchanan
                  Street, Albany, CA 94710, USA
  COMMENT         On Jan 5, 2001 this sequence version replaced gi:4713943.
                  This sequence is of BAC T8K14 from Arabidopsis thaliana chromosome
                  1.  In order to facilitate the joining of overlapping clones in the
                  future for creation of larger contigs, we provide overlap between
                  overlapping submitted clones.  The 3' end of this sequence overlaps
                  by 2000 bp the 5' end of the sequence of the YAC YUP8H12R.
  FEATURES             Location/Qualifiers
      source           1. .80374
                       /organism="Arabidopsis thaliana"
                       /cultivar="Columbia"
                       /db_xref="taxon:3702"
                       /chromosome="1"
                       /clone="T8K14"
      gene             448. .5132
                       /gene="T8K14.1"
      CDS              join(448. .3246,3320. .3391,3652. .3804,3894. .4057,
                       4156. .4303,4397. .4519,4737. .4824,4933. .5132)
                       /gene="T8K14.1"
                       /note="Is a member of the PF|00069 Eukaryotic protein
                       kinase family.  ESTs gb|T46484, gb|AF066875 and gb|N96237
                       come from this gene."
                       /codon_start=1
                       /evidence=not_experimental
                       /protein_id="AAD30219.1"
                       /db_xref="GI:4835752"
                       /translation="MDKARHQQLFQHSMEPGYRNETVPQPFMPDQTGSASANMRPPNS
                       NGSDVKAVHNFSIQTGEEFSLEFMRDRVIPQRSSNPNGAGDMNYNTGYMELRGLIGIS
                       HTGSECASDVSRFSTVENGTSDIERTNSSLHEFGNKLNHVQSAPQALLSKDSSVGNLH
                       GYKNTSSSASGSVTAKVKILCSFGGKILPRPGDSKLRYVGGETHIISIRKDISWQELR
                       QKILEIYYQTRVVKYQLPGEDLDALVSVSSEEDLQNMLEEYNEMENRGGSQKLRMFLF
                       SISDMDDALLGVNKNDGDSEFQYVVAVNGMDIGSGKNSTLLGLDSSSANNLAELDVRN
                       TEGINTIAGDVVGVGASQLMVNGFQQTSAQQSESIPPSSSLHYSQSIPLNAAYQLQQS
                       VPPSSALHYPQSITPGSSLQYPQSITPGSSYQYPQSIIPGSASSYGIYPQYYGHVVQH
                       GERERFPLYPDHSSNYSAIGETTSSIPIQGHVSQQGGWAEGYPYPGSTPKSTQALAEE
```

```
                          QKVSSDMKIREEVEPENRKTPGNDHQNPPQIDDVEVRNHNQVREMAVATTPPSQDAHL
                          LPPSRDPRQNTTAKPATYRDAVITGQVPLSGIEDQLSTSSSTYAPVHSDSESNLIDLN
                          YPEPEQSSQRVYCSERIPREQLELLNRLSKSDNSLSSQFVTSESPANTAQQDSGKEAV
                          GKSHDEFKTVNDDANHHTHKDVETIFEKVGVSDETLESEPLHKIVNPDDANKNRVVNG
                          ADTEIGVSNLSHVNAAMSHVIPEEQASLQGDILIDINDRFPRDFLSEIFSQAISEDTS
                          TVRPYPHDGAAVSMNVQNHDRKNWSYFQQLAEDQFIQRDVVLDQADSRIPSDRKDGGE
                          SSRLPYVSPLSRDGISTNLANPQLTLGQDYGGNFSEKDGGGTGSIPPALENEQMKVTE
                          SEEFGAMVENLRTPDSEPKDEKTETRHAALPPLGSEFDYSGLQIIKNEDLEELRELGS
                          GTFGTVYHGKWRGSDVAIKRIKKSCFAGRSSEQERLTGEFWGEAEILSKLHHPNVVAF
                          YGVVKDGPGGTLATVTEYMVDGSLRHVLVRKDRHLDRRKRLIIAMDAAFGMEYLHSKN
                          TVHFDLKCDNLLVNLKDPSRPICKVGDFGLSKIKRNTLVSGGVRGTLPWMAPELLNGS
                          SSKVSEKVDVFSFGIVLWEILTGEEPYANMHYGAIIGGIVNNTLRPTIPGFCDDEWRT
                          LMEECWAPNPMARPSFTEIAGRLRVMSSAATSTQSKPSAHRASK"
          gene            complement(5680. .11012)
                          /gene="T8K14.2"
          CDS             complement(join(5680. .5757,5848. .5928,6021. .6128,
                          6221. .6295,6397. .6516,6700. .6876,6987. .7103,7197.
.7316,

                          7406. .7443,7567. .7787,7887. .7999,8095. .8262,8367.
.8507,

                          8611. .8853,8933. .9010,9243. .9520,9737. .9963,10170.
.10354,

                          10584. .11012))
                          /gene="T8K14.2"
                          /note="Is a member of PF|00004 ATPases associated with
                          various cellular activities (AAA) family.  ESTs gb|T43031,
                          gb|R64750, gb|AA394742 and gb|AI100347 come from this
                          gene."
                          /codon_start=1
                          /evidence=not_experimental
                          /protein_id="AAD30220.1"
                          /db_xref="GI:4835753"
                          /translation="MEIAISYKPNPLISSSTQLLKRSKSFGLVRFPAKYGLGATRKKQ
                          LFRVYASESSSGSSSNSDGGFSWVRLAQSIRLGAERIGEKIGESVKTEIGFDSEEASG
                          RVNEYVARVKDSVHKGHHELTRFKNETVPSFIDWNKWEHWKDIRNWDGKRVAALFIYA
                          FALLLSCQRVYVAIQAPRVERERRELTESFMEALIPEPSPGNIEKFKRNMWRKATPKG
                          LKLKRFIEAPDGTLVHDSSYVGENAWDDDLETTEGSLKKIIGRNARIQTEAKKKLSQD
                          LGVSGEIGDSVGNWRERLATWKEMLEREKLSEQLNSSAAKYVVEFDMKEVEKSLREDV
                          IGRTSETEGTRALWISKRWWRYRPKLPYTYFLQKLDSSEVAAVVFTEDLKRLYVTMKE
                          GFPLEYIVDIPLDPYLFETICNAGVEVDLLQKRQIHYFMKVFIALLPGILILWFIRES
                          AMLLLITSKRFLYKKYNQLFDMAYAENFILPVGDVSETKSMYKEVVLGGDVWDLLDEL
                          MIYMGNPMQYYEKDVAFVRGVLLSGPPGTGKTLFARTLAKESGLPFVFASGAEFTDSE
                          KSGAAKINEMFSIARRNAPAFVFVDEIDAIAGRHARKDPRRRATFEALIAQLDGEKEK
                          TGIDRFSLRQAVIFICATNRPDELDLEFVRSGRIDRRLYIGLPDAKQRVQIFGVHSAG
                          KNLAEDIDFGKANIRNLVNEAAIMSVRKGRSYIYQQDIVDVLDKQLLEGMGVLLTEEE
                          QQKCEQSVSYEKKRLLAVHEAGHIVLAHLFPRFDWHAFSQLLPGGKETAVSVFYPRED
                          MVDQGYTTFGYMKMQMVVAHGGRCAERVVFGDNVTDGGKDDLEKITKIAREMVISPQS
                          ARLGLTQLVKKIGMVDLPDNPDGELIKYRWDHPHVMPAEMSVEVSELFTRELTRYIEE
                          TEELAMNALRANRHILDLITRELLEKSRITGLEVEEKMKDLSPLMFEDFVKPFQINPD
                          DEELLPHKDRVSYQPVDLRAAPLHRS"
          gene            11693. .13641
                          /gene="T8K14.3"
          CDS             join(11693. .11887,12188. .12265,12354. .12614,12703.
.13020,

                          13129. .13287,13447. .13641)
                          /gene="T8K14.3"
                          /note="Is a member of the PF|00162 Phosphoglycerate kinase
```

family. ESTs gb|N38721, gb|T22178, gb|R90345, gb|R90715,
gb|T21140, gb|T46295, gb|H37082, gb|T46076, gb|N37132,
gb|AA597649, gb|AI100648 and gb|Z48462 come from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30221.1"
/db_xref="GI:4835754"
/translation="MATKRSVGTLKEADLKGKSVFVRVDLNVPLDDNSITDDTRIRA
AVPTIKYLMGNGSRVVLCSHLGRPKGVTPKYSLKPLVPRLSELLGVEVVMANDSIGEE
VQKLVAGLPEGGVLLLENVRFYAEEEKNDPEFAKKLAALADVYVNDAFGTAHRAHAST
EGVAKFLKPSVAGFLMQKELDYLVGAVANPKKPFAAIVGGSKVSTKIGVIESLLNTVD
ILLLGGGMIFTFYKAQGLSVGSSLVEEDKLDLAKSLMEKAKAKGVSLLLPTDVVIADK
FAPDANSKIVPATAIPDGWMGLDIGPDSIKTFSEALDTTKTIIWNGPMGVFEFDKFAA
GTEAVAKQLAELSGKGVTTIIGGGDSVAAVEKVGLADKMSHISTGGGASLELLEGKPL
PGVLALDEA"

gene	15312. .17654
	/gene="T8K14.4"
CDS	15312. .17654

/gene="T8K14.4"
/note="Contains similarity to gi|2827663 F18F4.190
membrane-associated salt-inducible-like protein from
Arabidopsis thaliana BAC gb|AL021637."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30222.1"
/db_xref="GI:4835755"
/translation="MKPQMFFRSVIQFYSKPSWMQRSYSSGNAEFNISGEVISILAKK
KPIEPALEPLVPFLSKNIITSVIKDEVNRQLGFRFFIWASRRERLRSRESFGLVIDML
SEDNGCDLYWQTLEELKSGGVSVDSYCFCVLISAYAKMGMAEKAVESFGRMKEFDCRP
DVFTYNVILRVMMREEVFFMLAFAVYNEMLKCNCSPNLYTFGILMDGLYKKGRTSDAQ
KMFDDMTGRGISPNRVTYTILISGLCQRGSADDARKLFYEMQTSGNYPDSVAHNALLD
GFCKLGRMVEAFELLRLFEKDGFVLGLRGYSSLIDGLFRARRYTQAFELYANMLKKNI
KPDIILYTILIQGLSKAGKIEDALKLLSSMPSKGISPDTYCYNAVIKALCGRGLLEEG
RSLQLEMSETESFPDACTHTILICSMCRNGLVREAEEIFTEIEKSGCSPSVATFNALI
DGLCKSGELKEARLLLHKMEVGRPASLFLRLSHSGNRSFDTMVESGSILKAYRDLAHF
ADTGSSPDIVSYNVLINGFCRAGDIDGALKLLNVLQLKGLSPDSVTYNTLINGLHRVG
REEEAFKLFYAKDDFRHSPAVYRSLMTWSCRKRKVLVAFNLWMKYLKKISCLDDETAN
EIEQCFKEGETERALRRLIELDTRKDELTLGPYTIWLIGLCQSGRFHEALMVFSVLRE
KKILVTPPSCVKLIHGLCKREQLDAAIEVFLYTLDNNFKLMPRVCNYLLSSLLESTEK
MEIVSQLTNRMERAGYNVDSMLRFEILKYHRHRKQVLIDL"

gene	18900. .21756
	/gene="T8K14.5"
CDS	join(18900. .18995,19434. .19507,19606. .19673,19753.
.19791,	

19928. .20028,20168. .20280,20369. .20468,20547. .20693,
20774. .20834,20925. .21022,21110. .21252,21361. .21444,
21529. .21618,21702. .21756)
/gene="T8K14.5"
/note="Is a member of the PF|00044 glyceraldehyde
3-phosphate dehydrogenase family. ESTs gb|T43985,
gb|N38667, gb|N65037, gb|AA713069 and gb|AI099548 come
from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30223.1"
/db_xref="GI:4835756"

```
                        /translation="MAFSSLLRSAASYTVAAPRPDFFSSPASDHSKVLSSLGFSRNLK
                        PSRFSSGISSSLQNGNARSVQPIKATATEVPSAVRRSSSSGKTKVGINGFGRIGRLVL
                        RIATSRDDIEVVAVNDPFIDAKYMAYMLKYDSTHGNFKGSINVIDDSTLEINGKKVNV
                        VSKRDPSEIPWADLGADYVVESSGVFTTLSKAASHLKGGAKKVIISAPSADAPMFVVG
                        VNEHTYQPNMDIVSNASCTTNCLAPLAKVVHEEFGILEGLMTTVHATTATQKTVDGPS
                        MKDWRGGRGASQNIIPSSTGAAKAVGKVLPELNGKLTGMAFRVPTSNVSVVDLTCRLE
                        KGASYEDVKAAIKHASEGPLKGILGYTDEDVVSNDFVGDSRSSIFDANAGIGLSKSFV
                        KLVSWYDNEWGYSNRVLDLIEHMALVAASH"
        gene            23473. .26525
                        /gene="T8K14.6"
      CDS               join(23473. .23665,24252. .24364,24454. .24699,24777.
.24857,
                        24955. .25185,25275. .25613,26424. .26525)
                        /gene="T8K14.6"
                        /note="EST gb|AA404917 comes from this gene."
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAD30224.1"
                        /db_xref="GI:4835757"

    Query Match          22.0%;  Score 93.6;  DB 8;  Length 80374;
    Best Local Similarity  64.1%;  Pred. No. 5e-13;
       Matches  141;  Conservative   0;  Mismatches   79;  Indels   0;  Gaps   0;


Qy     49 atttcaagttcaagcaagagctctggatggtcattagcatgtcctctgttgcggtcgtga 108
          || | |      || | ||    |||||| |      |   ||| | || | | | |||| |
Db  24965 ATATGAGTAGCACCGAGGAAAAATGGATGATTGGAATAATGGCTTCAGCTACAGTCGTCA 25024


Qy    109 agttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgccc 168
          ||||  | ||||||||| |||||| | |  ||| |||| || || || |||||||| || |
Db  25025 AGTTTCTGCTCATGCTTTACTGCAGGAGTTTCCAGAACGAAATTGTCAGGGCCTATGCAC 25084


Qy    169 aggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctg 228
          | || ||  |||| | | ||||| || || |||| |   | | | | | ||  | |||||
Db  25085 AAGATCACCTCTTTGATGTTATCACCAATTCAGTCGGTTTAGCAACCGCTGTTTTAGCTG 25144


Qy    229 tccggtacaaatggtggatggaccctgttggcgccatact 268
          |     | | | ||||||||| || ||    ||| || ||||||
Db  25145 TAAAATTCTACTGGTGGATTGATCCCTCTGGGGCTATACT 25184



RESULT    5
CELR02F11/c
LOCUS       CELR02F11   33270 bp    DNA              INV        07-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid R02F11.
ACCESSION   AF016439
VERSION     AF016439.1  GI:2315347
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
   ORGANISM  Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 33270)
   AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
             Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
             Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
             Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
```

```
             Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
             Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
             O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
             Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
             Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
             Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
             Wilkinson-Sproat,J. and Wohldman,P.
  TITLE      2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
             elegans
  JOURNAL    Nature 368 (6466), 32-38 (1994)
  MEDLINE    94150718
REFERENCE   2  (bases 1 to 33270)
  AUTHORS    Davidson,S. and Wohldmann,P.
  TITLE      The sequence of C. elegans cosmid R02F11
  JOURNAL    Unpublished (1997)
REFERENCE   3  (bases 1 to 33270)
  AUTHORS    Waterston,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (30-JUL-1997)
```
```
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone


                NEIGHBORING COSMID INFORMATION:


            The 5' end of this cosmid lies in a gap;3' cosmid is C37H5, 1301 bp
            overlap.  Actual start of this cosmid is at base position 1 of
            CELR02F11; actual end is at 11162 of CELC37H5.  This cosmid lies in
            an unanchored cluster, the orientation of which is unknown.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..33270
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
```

```
                             /chromosome="V"
                             /clone="R02F11"
        gene                 340. .9745
                             /gene="R02F11.2"
        CDS                  join(340. .470,814. .1029,1073. .1217,2116. .2417,2486.
.2695,

                             2739. .2872,3558. .3608,3658. .3725,3772. .3937,3993.
.4314,

                             5822. .6005,7672. .7733,8045. .8197,8494. .8620,8669.
.8781,

                             8829. .8902,9591. .9745)
                             /gene="R02F11.2"
                             /note="contains similarity to a lectin C-type domain"
                             /codon_start=1
                             /evidence=not_experimental
                             /protein_id="AAB65898.1"
                             /db_xref="GI:2315349"
                             /translation="MRFIFFAIIFFEAVLAFDVPLFSTSSSKGKRNGSIVQDPVAPIY
                             KFGITYYWAGHYKSKHGIQNKCVFLKSDPDWPFPESTFSNGTAVYGVEFGCKKDDLCR
                             KYHCDRPFQTGFIWCAFLILSSVISVGIIVRYIIMQRQNELRPTRDTELGSIYIPVLM
                             INGDSSSGAQPSYFVSFGKIGISATCRFLNLSEIFNFGKLPICRKFQFRQFADLPEQV
                             VCRLPTTRVTNLPIAYLNIFLMQVVIGIGTKQAPFLRMELLLSNVLSYHGIAFYWSGA
                             YIPSPERPETCIILASHPEWPFDGNVFQNGTIPTAALFGCHERSMCSGTRCIEPYTHF
                             NVVTYTLLGIVLFLLVCLLGGEPKNKEKSKRRVSRHSTLDNIFLPLLFFIGFAAAGNY
                             GYSSSGNGGGRPRPGNGGGGGGVRPSNSGCDAGWRRFNRPSGGWCVKVFGARMNQADA
                             QIQCQSHGATLSGLQNSEEAQQISNLALSVISANSGSVWIGTRRTAACMKQWLNTNGC
                             TRTNAFYWTDGSATGIAGFVWDTLQPDNEKLSQSCAVLLASRSTVTWGGKFWQPAKMD
                             DNNCLFDLEGKHPRSVSLNALLFIWLLRFLITLCIPERRKPSTTSGKSTRQSQDGGSV
                             ETTNLLTPIPLTPFYVESVRPTVHWPSISNTQISGPLDVGETFLAGSEYEQSLIGHLL
                             NSSAMVTLNDDITTLLASTNMTYTLASPDEPFVFVERNYFWSQKDFNEYADGENNTIF
                             SYVCVYNASEDEGLFSQIYFPSGDRIQEVVFGCEVSKECCGMKCCGDDVLINIIIVGV
                             ISLALLLLLLCNILIGFKKRREKSRETNLKATYQATDTTNGAIDNDIISADQITYDTR
                             HPATNRGSPANPTA"
        gene                 14022. .15284
                             /gene="R02F11.1"
        CDS                  join(14022. .14103,14148. .14315,15004. .15284)
                             /gene="R02F11.1"
                             /note="similar to granulins; coded for by C. elegans cDNA
                             CEESU12R; coded for by C. elegans cDNA yk121b8.5"
                             /codon_start=1
                             /protein_id="AAB65897.1"
                             /db_xref="GI:2315348"
                             /translation="MKQLAISFCIILPLACFWVLNAEDAKEAPEIREANSTVFRQKRQ
                             FGCPSNCYSSCSSSSQCQRYSVASVCVQGCCCPGNNNLDTACSGGPAVAACLGGLCGQ
                             GFFCSSNNYCCRCQSGNSTGPCVNQVCPTGFMCNTNNYCCPLGSGGVLGSCVNGVCPT
                             GYTCGAGNLCYLSSGK"
        gene                 complement(15794. .18813)
                             /gene="R02F11.3"
        CDS                  complement(join(15794. .15955,16141. .16559,16708. .16833,
                             16926. .17313,18313. .18438,18505. .18813))
                             /gene="R02F11.3"
                             /codon_start=1
                             /evidence=not_experimental
                             /protein_id="AAB65899.1"
                             /db_xref="GI:2315350"
                             /translation="MTKTEATFSLLDEEEERDEVGPSTVLVPGRPSLRSISLSQAPSA
                             PRDLENGNGQTSCAELHKLTAGSTTSLSSQSKNAKKVNKFYKKQNELLENFKNDSEQI
```

```
                        EQFNRTRRRTTSKEEDDDADVIAAIPPPIPEEKAVVAPLVKHIKPMERTDTEEKSVDE
                        KKDDESNTAARMANITLAVNFLLMIAKVVASVLSGSMSIISSMVDSVVDITSGLVISL
                        SERMIKKRDPYLYPRGRTRLEPLSLILISVIMGMASIQLIIASVRGIHDGIQFHLYGA
                        LLIFEPISVFELILSDWLETELISLAVLSFGGNLQYIRPQRIGEEPKLNVTITSVVIM
                        VSTVLVKLSLYLFCKRYKEPSVNVLAMDHRNDCISNTVALICAWLGTKYSYYFDPAGA
                        IVVSMYILYTWVQTGREHLAKLSGKTAEPEFINRIIKVCLDHDARISHIDTVYVYHFG
                        SKFLVEVHIVLDENMILKESHDISETLQSNIESLPEVERAFVHTDYDYDHHPHDEHKI
                        V"
          gene          complement(23977. .31536)
                        /gene="R02F11.4"
          CDS           complement(join(23977. .24039,25118. .25722,27089. .27586,
                        28991. .29114,29171. .29408,30288. .30343,30401. .30578,
                        31406. .31536))
                        /gene="R02F11.4"
                        /note="weak similarity to Listeria monocytogenes
                        internalin A (SP:P25146) and lumincans; coded for by C.
                        elegans cDNA yk109c2.5; coded for by C. elegans cDNA
                        yk109c2.3"
                        /codon_start=1
                        /protein_id="AAB65900.1"
                        /db_xref="GI:2315351"
                        /translation="MSEEDEPKKAAADAADGEDDDKENLDDSRANSSLANSTFVTSGR
                        IQNEKADVEVDGTLDLANLSLTNLEKSFAAEYSEVKHLKISGNCFQKFTYIKLFPKCQ
                        IIDARDCQINKFIADFNYNLLELHLARNQLKETNQLGRFENLKILDLSNNLIEPPVSF
                        SLKNLEILNLSGNFLNEIPDLSKCVALQTISLADNKISDLTTITKLICPTNLKNLDIS
                        SNSIEDLSQFSVLSTFKKLEEFVVAGNPSITSVLDSDLFDYRSYIFACCSEQLHTIDG
                        QKIEDQVQTEGEWLALQGSIKKIGPGNHDALCQQIASHFPDSGPPTPAQKSCHKALEK
                        RRSMKVPEKHLEEFSEDTDRTENSVYSPFREWNGKIGALMTPEASGSSRRIPVNKNLR
                        MCSPPEARKNKTFKFQNETSPNRYPENFRPLGSTRSISTETVICSSRTEVSFTVDGRS
                        ESTPLPRIDCAPLERKEQEPEQRHTPSVADVTYVGEESDGELKQKVAYLEKSVDELKI
                        QNENLTAINDRLVDTLEEFKAEQAEMWKAIRNMIPTPQNVISNFLVENEDGHHVHQVR
                        WDMPAVKEFRIFVDGNQCGQIKGKNNSARITDLTSNEPHTVQIQAVGNNGLHGEISKK
                        LHIVPK"
BASE COUNT      10624 a    5767 c    5928 g   10951 t
ORIGIN
```

```
     Query Match               21.0%;   Score 89.4;   DB 3;   Length 33270;
     Best Local Similarity     52.6%;   Pred. No. 5.8e-12;
     Matches  195; Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy      48 gatttcaagttcaagcaagagctctggatggtcattagcatgtcctctgttgcggtcgtg 107
           ||    |||| ||||         ||        || | | |||| || |   | ||||
Db   16548 GAACCCAAGCTCAACGTCACCATCACCTCTGTAGTCATCATGGTGTCAACAGTTCTCGTC 16489

Qy     108 aagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcc 167
           ||| | | |||| || | |||   | | |||| |         || | |   | |||
Db   16488 AAGCTGTCCCTCTACCTATTCTGTAAACGATACAAGGAACCATCGGTCAACGTGCTCGCA 16429

Qy     168 caggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgct 227
             |||||||  |  |||      ||| | ||| | |||| |||| ||| ||    |||||
Db   16428 ATGGACCATCGCAACGATTGCATCTCCAACACGGTCGCCCTGATCTGTGCCTGGCTCGGC 16369

Qy     228 gtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgatc 287
            |  ||||     |  | |  | |||||| | || || ||  || |   | ||||||   |
Db   16368 ACCAAGTACTCGTACTACTTTGACCCAGCCGGTGCTATTGTGGTTTCTATGTACATTTTG 16309
```

```
Qy    288 acgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccg 347.
              || |||| || ||| | | |||| | | | | ||| || || | ||  |
Db  16308 TATACCTGGGTGCAAACTGGACGGGAGCATTTGGCAAAGCTGTCGGGTAAAACTGCAGAG 16249

Qy    348 gcagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcac 407
              |||||| | | | | | ||| | | | || | | || ||| ||| ||
Db  16248 CCAGAGTTCATTAATAGGATCATCAAAGTCTGCTTGGATCATGATGCTCGGATTTCACAT 16189

Qy    408 atcgacacggt 418
              || || |||||
Db  16188 ATTGATACGGT 16178
```

RESULT   6
AP003231
```
LOCUS       AP003231   138108 bp    DNA              PLN       05-JUL-2001
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0031D11,
            complete sequence.
ACCESSION   AP003231
VERSION     AP003231.2  GI:14624986
KEYWORDS    HTG.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0031D11.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 138108)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0031D11
  JOURNAL   Published Only in Database (2001) In press
REFERENCE   2  (bases 1 to 138108)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jul 5, 2001 this sequence version replaced gi:13027261.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
FEATURES             Location/Qualifiers
     source          1. .138108
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0031D11"
BASE COUNT     39143 a   30307 c   30396 g   38262 t
ORIGIN


   Query Match              19.2%;  Score 82;  DB 8;  Length 138108;
   Best Local Similarity    69.1%;  Pred. No. 3.3e-10;
   Matches 112; Conservative   0;  Mismatches  50;  Indels   0;  Gaps   0;

Qy    265 tactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtaggca 324
```

```
          |  |  |  |  ||   |  ||||||||||   ||  |||  ||    |||||||||  ||||||||    |
Db 128084 TGCAGCTAGCAATCTACACGATCAGAACATGGTCGATGACGGTGCTCGAGAACGTCCACT 128143

Qy    325 cactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatctgga 384
          |  |||  |  ||   ||||  ||   ||  |  ||  |||||    ||||||  ||  |||  |    |||||
Db 128144 CCCTGGTCGGGCAGTCAGCCTCGCCGGAATACCTTCAGAAGCTGACCTACCTATGCTGGA 128203

Qy    385 accaccatgaggagatccagcacatcgacacggtgcgagcct 426
          |||||||  |||    |    ||||||  |||||||||||||  ||  |
Db 128204 ACCACCACAAGGCCGTGAGGCACATAGACACGGTGCGGGCGT 128245



RESULT   7
AC004218/c
LOCUS       AC004218    86950 bp    DNA              PLN        05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 212 of 255 of the
            complete sequence. Sequence from clones F12L6.
ACCESSION   AC004218 AE002093
VERSION     AC004218.2  GI:6598409
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 86950)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
    TITLE   Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 86950)
  AUTHORS   Lin,X.
    TITLE   Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:3355463.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
```

databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

                Address all correspondence to: at@tigr.org.
FEATURES   Location/Qualifiers
    source   1. .86950
                        /organism="Arabidopsis thaliana"
                        /cultivar="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="II"
    misc_feature   <1. .>86950
                        /note="Sequence from clone F12L6"
    mRNA   complement(<2699. .>5146)
                        /gene="At2g39360"
    gene   complement(<2699. .>5146)
                        /gene="At2g39360"
                        /note="F12L6.2; contains a protein kinase domain profile (PDOC00100)"
    CDS   complement(2699. .5146)
                        /gene="At2g39360"
                        /codon_start=1
                        /product="putative protein kinase"
                        /protein_id="AAC27827.1"
                        /db_xref="GI:3355465"
                        /translation="MINLKLFLELKLCFLITLLCSSHISSVSDTFFINCGSPTNVTVN
                        NRTFVSDNNLVQGFSVGTTDSNSGDESTLFQTARVFSDESSSTYRFPIEEHGWFLIRI
                        YFLPLVSASQDLTTARFSVSAQNFTLIREYKPSTTSVVREYILNVTTDSLLLQFLPRT
                        GSVSFINALEVLRLPETLIPEDAKLIGTQKDLKLSSHAMETVSRVNMGNLSVSRDQDK
                        LWRQWDSDSAYKAHFGTPVMNLKAVNFSAGGITDDIAPVYVYGTATRLNSDLDPNTNA
                        NLTWTFKVEPGFDYFVRFHFCNIIVDPFGFERQIRFDIFVNSEKVRTIDMTEVLNGTF
                        GAPFFVDAVMRKAKSREGFLNLSIGLVMDVSSYPVSFINGFEISKLSNDKRSLDAFDA
                        ILPDGSSSNKSSNTSVGLIAGLSAALCVALVFGVVVSWWCIRKRRRRNRQMQTVHSRG
                        DDHQIKKNETGESLIFSSSKIGYRYPLALIKEATDDFDESLVIGVGGFGKVYKGVLRD
                        KTEVAVKRGAPQSRQGLAEFKTEVEMLTQFRHRHLVSLIGYCDENSEMIIVYEYMEKG

```
                         TLKDHLYDLDDKPRLSWRQRLEICVGAARGLHYLHTGSTRAIIHRDVKSANILLDDNF
                         MAKVADFGLSKTGPDLDQTHVSTAVKGSFGYLDPEYLTRQQLTEKSDVYSFGVVMLEV
                         VCGRPVIDPSLPREKVNLIEWAMKLVKKGKLEDIIDPFLVGKVKLEEVKKYCEVTEKC
                         LSQNGIERPAMGDLLWNLEFMLQVQAKDEKAAMVDDKPEASVVGSTMQFSVNGVGDIA
                         GVSMSKVFAQMVREETR"
     repeat_region      complement(7060. .7174)
                         /rpt_family="POLY_A"
     repeat_region      7725. .7846
                         /rpt_family="(TAAA)n"
     repeat_region      complement(8238. .8272)
                         /rpt_family="(CATA)n"
     repeat_region      8666. .8705
                         /rpt_family="(TA)n"
     mRNA               complement(9237. .>10373)
                         /gene="At2g39370"
     gene               complement(9237. .>10373)
                         /gene="At2g39370"
                         /note="F12L6.3"
     CDS                complement(9387. .10373)
                         /gene="At2g39370"
                         /note="unknown protein"
                         /codon_start=1
                         /protein_id="AAC27828.1"
                         /db_xref="GI:3355466"
                         /translation="MAAYLERCDSVEEDYIDMEVTSFTNLVRKTLSNNYPREFEFQMS
                         HLCPLEIDKTTSPADELFYKGKLLPLHLPPRLQMVQKILEDYTFDDEFYSTPLATGTV
                         TTPVTSNTPFESCTVSPAESCQVSKELNPEDYFLEYSDSLEEDDEKKKSWTTKLRLMK
                         QSSLGTKIKASRAYLRSFFGKTSCSDESSCASSAARVADEDSVLRYSRVKPFGQIKTE
                         RPKKQSNGSVSGSHRRSFSVSMRRQAAKSSNNKSSNSLGFRPLQFLKRSTSSSSEIEN
                         SIQGAILHCKQSQQQKQKQKQYSTVNEVGFCSLSASRIAARDDQEWAQMFRG"
     repeat_region      11459. .11485
                         /rpt_family="POLY_A"
     mRNA               complement(<12203. .>14116)
                         /gene="At2g39380"
     gene               complement(<12203. .>14116)
                         /gene="At2g39380"
                         /note="F12L6.4; predicted by grail"
     CDS                complement(12203. .14116)
                         /gene="At2g39380"
                         /note="hypothetical protein"
                         /codon_start=1
                         /protein_id="AAC27829.1"
                         /db_xref="GI:3355467"
                         /translation="MSKKPKSVHFSTSSPKSFLSSFPSFTSLPASPLNQTFSQSMMEE
                         TVEAAESIIKKWDPNSPSYTKIISLFSHSRREAKEFIRCIRDLRRAMHFLISQHSKSA
                         KLVLAQHLMQIAMARLEKEFFQILSSNRDQLDPESVSGHSSISSNSEFEDVMQSSDDE
                         EEDELKKAGETITKVEKAAALVMSDLKVIAETMISCGYGKECIKSYKLIRKSIVDEGL
                         HLLGIEKCKISRFNRMDWDVLEHMIKNWIKAAKIGVITLLRGEKLLCDHVFSASSTIR
                         ESCFYEIVNEAGINLFKFPELVAEKKPSPERIFRLMDLYAAISDLRPDIELIFHFDSV
                         AAVKTLVLSSLKKLKDSIYTSLMEFESTIQKDSSKALTAGGGIHKLTRSTMSFISSLS
                         EYSRVLSEILAEHPLKKNTRMLESYFTAPILEDEHNNHAVSVHLAWLILIFLCKLDIK
                         AESYKDVSLSYLFLVNNIQFVVDTVRSTHLRNLLGDDWLTKHEAKLRSYAANYEIAAW
                         ANVYISLPEKTSSRLSPEEAKTHFKRFHAVFEEAYMKQSSCVITDAKLRNELKVSIAK
                         KIVPEYREFYGKYLPTLSKERNIEMLVSFKPDNLENYLSDLFHGTPILSGSSSSSSSL
                         SSSSCISLGCVRN"
     repeat_region      12232. .12268
                         /rpt_family="(GAA)n"
```

```
       mRNA                complement(join(<15910. .15984,16090. .16246,16631. .16766,
                           16866. .16939))
                           /gene="At2g39390"
       gene                complement(<15910. .16939)
                           /gene="At2g39390"
                           /note="F12L6.5"
       CDS                 complement(join(15910. .15984,16090. .16246,16631. .16766,
                           16866. .16869))
                           /gene="At2g39390"
                           /codon_start=1
                           /product="60S ribosomal protein L35"
                           /protein_id="AAC27830.1"
                           /db_xref="GI:3355468"
                           /translation="MARIKVHELREKSKADLSGQLKEFKAELALLRVAKVTGGAPNKL
                           SKIKVVRKSIAQVLTVISQKQKSALREAYKNKKLLPLDLRPKKTRAIRRRLTKHQASL
                           KTEREKKKEMYFPIRKYAIKV"
       mRNA                join(<17968. .18130,18223. .18369,18454. .18584,19041.
.19178,
                           19270. .19368,19482. .>19757)
                           /gene="At2g39400"
       gene                <17968. .>19757
                           /gene="At2g39400"
                           /note="F12L6.6"
       CDS                 join(17968. .18130,18223. .18369,18454. .18584,19041.
.19178,
                           19270. .19368,19482. .19757)
                           /gene="At2g39400"
                           /codon_start=1
                           /product="putative phospholipase"
                           /protein_id="AAC27831.1"
                           /db_xref="GI:3355469"
                           /translation="MQVCVFFLLAGFCVELTRHEVVHLCMETSEARTKGIALPLPWLR
                           YGVQHHHEQFSAATRLANAGFAVYGMDYEGHGKSEGLNGYISNFDDLVDDVSNHYSTI
                           CEREENKGKMRFLLGESMGGAVVLLLARKKPDFWDGAVLVAPMCKLADEIKPHPVVIS
                           ILIKLAKFIPTWKIVPGNDIIDIAIKEPHIRNQVRENKYCYKGRPRLNTAYQLLLVSL
                           DLEKNLHQVSIPFIVLHGEDDKVTDKSISKMLYEVASSSDKTFKLYPKMWHALLYGET
                           NENSEIVFGDIINWLEDRATDSNGGLESQLKHKHDGFLKHK"
       mRNA                join(<20854. .20871,20986. .21112,21306. .21452,21671.
.21801,
                           22519. .22656,22777. .22875,23049. .>23324)
                           /gene="At2g39410"
       gene                <20854. .>23324
                           /gene="At2g39410"
                           /note="F12L6.7"
       CDS                 join(20854. .20871,20986. .21112,21306. .21452,21671.
.21801,
                           22519. .22656,22777. .22875,23049. .23324)

   Query Match             19.2%;   Score 81.8;   DB 8;   Length 86950;
   Best Local Similarity   62.4%;   Pred. No. 3.9e-10;
   Matches  128;  Conservative   0;  Mismatches   77;  Indels   0;  Gaps   0;

Qy     61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
           |||||||   ||| | ||     | |||     |||||| | | || || | | | ||
Db  37484 AGCAAGAGAGTTGGGTAGTTGGGATCATGCTTTCTGTTACATTGGTCAAACTGCTTCTGG 37425

Qy    121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
```

```
              |  ||  ||||||  ||  |  ||||   ||  ||||||||||  |   ||  ||  ||  ||  ||  ||  |||||||
Db   37424  TTCTTTACTGCAGATCCTTCACTAACGAGATCGTTAAAGCTTATGCTCAAGATCATTTCT 37365


Qy     181  tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
              |||||||  |||||||||    |  ||  ||   |   ||    |  ||  ||         ||  ||
Db   37364  TCGACGTCATCACAAACATCATTGGACTCATTGCAGTAATCCTGGCCAATTACATTGATT 37305


Qy     241  ggtggatggaccctgttggcgccat 265
              |||||  ||  ||  |||||  ||  ||
Db   37304  ATTGGATTGATCCAGTTGGAGCTAT 37280



RESULT    8
ATT10K17/c
LOCUS       ATT10K17    109016 bp    DNA              PLN        20-JAN-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T10K17.
ACCESSION   AL132977
VERSION     AL132977.1  GI:6434223
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 109016)
  AUTHORS   Benes,V., Wurmbach,E., Drzonek,H., Ansorge,W., Mewes,H.W.,
            Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109016)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .109016
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
     misc_feature    1. .5127
                     /note="overlap to BAC F15B8, please refer to Acc no.
                     EMBL:AL049660 for analysis and annotation"
     gene            3113. .6400
                     /gene="T10K17.10"
     exon            complement(3113. .3385)
                     /number=1
     CDS             complement(join(3113. .3385,4086. .4157,4306. .4374,
                     5512. .5577,5669. .5782,5855. .6400))
                     /note="similarity to several hypothetical proteins -
```

```
                        Arabidopsis thaliana"
                        /codon_start=1
                        /product="putative protein"
                        /protein_id="CAB67608.1"
                        /db_xref="GI:6729523"
                        /translation="MDLTGGFGARSGGVGPCREPIGLESLHLGDEFRQLVTTLPPENP
                        GGSFTALLELPPTQAVELLHFTDSSSSQQAAVTGIGGEIPPPLHSFGGTLAFPSNSVL
                        MERAARFSVIATEQQNGNISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAI
                        ENQIPCPNQNNRNGKRKDFEKKGKSSTKKNKSSEENEKLPYVHVRARRGQATDSHSLA
                        ERARREKINARMKLLQELVPGCDKIQGTALVLDEIINHVQSLQRQVEMLSMRLAAVNP
                        RIDFNLDTILASENGSLMDGSFNAAPMQLAWPQQAIETEQSFHHRQLQQPPTQQWPFD
                        GLNQPVWGREEDQAHGNDNSNLMAVSENVMVASANLHPNQVKMEL"
      intron            complement(3386. .4085)
                        /number=1
      exon              complement(4086. .4157)
                        /number=2
      intron            complement(4158. .4305)
                        /number=2
      exon              complement(4306. .4374)
                        /number=3
      intron            complement(4375. .5511)
                        /number=3
      exon              complement(5512. .5577)
                        /number=4
      intron            complement(5578. .5668)
                        /number=4
      exon              complement(5669. .5782)
                        /number=5
      intron            complement(5783. .5854)
                        /number=5
      exon              complement(5855. .6400)
                        /number=6
      misc_feature      9838. .109016
                        /note="overlap to BAC F9D24"
      exon              11415. .11928
                        /gene="T10K17.20"
                        /number=1
      gene              11415. .12888
                        /gene="T10K17.20"
      CDS               join(11415. .11928,12048. .12130,12308. .12348,12502.
.12600,
                        12711. .12888)
                        /gene="T10K17.20"
                        /codon_start=1
                        /product="putative protein"
                        /protein_id="CAB67609.1"
                        /db_xref="GI:6729524"
                        /translation="MMICYSPITTCSRNAISIKRHLGSRLYGVVAHGSSKFSCYSLLS
                        GLSRRHYTGFRVSVSNRPSSWHDKGLFGSVLINRPTVAPKEKLEVSFLSPEANMKCSK
                        IESNMRNLYCYSRFAYTGVIVSLLVCYSSTSQSAYADSSRDKDANNVHHHSSDGKFHN
                        GKRVYTDYSIIAHGFCLRSGKLAPGEKMQRELADELRTRVADEFIQRRQETEWFVEGD
                        FDTYVRQIRDPHVWGGEPELFMASHVLQMPITVYMKDDKAGGLISIAEYGQEYGKDDP
                        IRVLYHGFGHYDALLLHESKASIPKSKL"
      intron            11929. .12047
                        /gene="T10K17.20"
                        /number=1
      exon              12048. .12130
```

```
                        /gene="T10K17.20"
                        /number=2
     intron             12131. .12307
                        /gene="T10K17.20"
                        /number=2
     exon               12308. .12348
                        /gene="T10K17.20"
                        /number=3
     intron             12349. .12501
                        /gene="T10K17.20"
                        /number=3
     exon               12502. .12600
                        /gene="T10K17.20"
                        /number=4
     intron             12601. .12710
                        /gene="T10K17.20"
                        /number=4
     exon               12711. .12888
                        /gene="T10K17.20"
                        /number=5
     gene               13360. .13903
                        /gene="T10K17.30"
     CDS                complement(join(13360. .13528,13803. .13903))
                        /note="ESTs matching in this region do not correspond to
                        any gene models eg.GB:AA651588; similarity to 60S
                        RIBOSOMAL PROTEIN L21 - Arabidopsis thaliana,
                        SWISSPROT:RL21_ARATH"
                        /codon_start=1
                        /product="putative protein"
                        /protein_id="CAB67610.1"
                        /db_xref="GI:6729525"
                        /translation="MHVNIPLMSPTTTVPDGIHTGSKQAYPSFTFGLKWVYQIGDKII
                        RKRIHVLVEHVQQSRCAVEFKLRKKKNDELKAASKARGETISTKR"
     exon               complement(13360. .13528)
                        /number=1
     intron             complement(13529. .13802)
                        /number=1
     exon               complement(13803. .13903)
                        /number=2
     gene               14857. .17399
                        /gene="T10K17.40"
     exon               14857. .16195
                        /gene="T10K17.40"
                        /number=1
     CDS                join(14857. .16195,16750. .17399)
                        /gene="T10K17.40"
                        /note="strong similarity to several receptor-like protein
                        kinases"
                        /codon_start=1
                        /product="receptor-like protein kinase"
                        /protein_id="CAB67611.1"
                        /db_xref="GI:6729526"
                        /translation="MQFLRLLTLLVSSYFFFFINFSSSLNPDGLSLLALKSAILRDPT
                        RVMTSWSESDPTPCHWPGIICTHGRVTSLVLSGRRLSGYIPSKLGLLDSLIKLDLARN
                        NFSKPVPTRLFNAVNLRYIDLSHNSISGPIPAQIQSLKNLTHIDFSSNLLNGSLPQSL
                        TQLGSLVGTLNLSYNSFSGEIPPSYGRFPVFVSLDLGHNNLTGKIPQIGSLLNQGPTA
                        FAGNSELCGFPLQKLCKDEGTNPKLVAPKPEGSQILPKKPNPSFIDKDGRKNKPITGS
```

```
                      VTVSLISGVSIVIGAVSISVWLIRRKLSSTVSTPEKNNTAAPLDDAADEEEKEGKFVV
                      MDEGFELELEDLLRASAYVVGKSRSGIVYRVVAGMGSGTVAATFTSSTVVAVRRLSDG
                      DATWRRKDFENEVEAISRVQHPNIVRLRAYYYAEDERLLITDYIRNGSLYSALHGGPS
                      NTLPSLSWPERLLIAQGTARGLMYIHEYSPRKYVHGNLKSTKILLDDELLPRISGFGL
                      TRLVSGYSKLIGSLSATRQSLDQTYLTSATTVTRITAPTVAYLAPEARASSGCKLSQK
                      CDVYSFGVVLMELLTGRLPNASSKNNGEELVRVVRNWVKEEKPLSEILDPEILNKGHA
                      DKQVIAAIHVALNCTEMDPEVRPRMRSVSESLGRIKSD"
        intron        16196. .16749
                      /gene="T10K17.40"
                      /number=1
        exon          16750. .17399
                      /gene="T10K17.40"
                      /number=2
        gene          17902. .18366
                      /gene="T10K17.50"
        exon          17902. .18366
                      /gene="T10K17.50"
                      /number=1
        CDS           17902. .18366
                      /gene="T10K17.50"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="CAB67612.1"
                      /db_xref="GI:6729527"
                      /translation="MASQRKLIMVVILSSLLMKVALSQYGVVMGKSIFKWEFFPMISV
                      YITNDIGGGLVLHSGCYTSRNGYRRIRDFFPGSMKIFAEFRKTYWGRTRYHCEFRFGD
                      ETQIHRFSLYKDIRDNIDKYQCRHCFWSIRRNGPCALNSHTGKYDICYAWDK"
        repeat_unit   18373. .18408
                      /note="36 bp TA tandem repeat"
        gene          19456. .19870
                      /gene="T10K17.60"
        exon          19456. .19543
                      /gene="T10K17.60"
                      /number=1
        CDS           join(19456. .19543,19686. .19870)
                      /gene="T10K17.60"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="CAB67613.1"
                      /db_xref="GI:6729528"
                      /translation="MASERKLIMVVILSSLLMKVALSQNGVVMGDYSFNSGKGFGEEL
                      AMFVNLDLGMKDKYVILLFIKITEIIFTNMSADTVFGPSEGMDLVP"
        intron        19544. .19685
                      /gene="T10K17.60"

  Query Match              13.0%;  Score 55.2;  DB 8;  Length 109016;
  Best Local Similarity    58.5%;  Pred. No. 0.0012;
  Matches   96;  Conservative   0;  Mismatches  68;  Indels   0;  Gaps   0;

Qy     169 aggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctg 228
               ||||  |||   ||| || || |·|||||   | |  ||  ||||  ||||  | || |||
Db   93640 AGGATCATCACTTTGATGTGGTAACAAATGTTCTTGGATTGGTTGCGGCCGTTCTTGCTA 93581

Qy     229 tccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgatca 288
                |   | ||||||| | || ||   ||| || || ||  | ||  | ||||| ||
Db   93580 ATGCTTTTTACTGGTGGCTCGATCCAACTGGTGCTATTCTCTTAGCCATCTACACTATTG 93521
```

```
Qy    289 cgacgtgggcgcgaacggtgctggagaacgtaggcacactgata 332
          |  ||| |  | || ||  |||| || | ||| | | |  ||
Db  93520 TCAATTGGTCTGGGACTGTTATGGAAAATGCAGGTAAATTTGTA 93477
```

RESULT   9
AF357202
LOCUS   AF357202   113193 bp   DNA   BCT   17-JUL-2001
DEFINITION Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION   AF357202.1 GI:14794889
KEYWORDS   .
SOURCE   Streptomyces nodosus.
  ORGANISM Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE   The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES   Location/Qualifiers
     source   1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene   complement(4. .1824)
                     /gene="amphG"
     CDS   complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene   complement(1805. .3628)
                     /gene="amphH"
     CDS   complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
```

```
                              /codon_start=1
                              /transl_table=11
                              /product="AmphH"
                              /protein_id="AAK73499.1"
                              /db_xref="GI:14794891"
                              /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                              LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                              LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                              FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                              GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                              FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                              EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                              GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                              ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                              ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                              DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
        gene                  3840..4874
                              /gene="amphDIII"
        CDS                   3840..4874
                              /gene="amphDIII"
                              /note="mycosamine biosynthesis protein"
                              /codon_start=1
                              /transl_table=11
                              /product="AmphDIII"
                              /protein_id="AAK73500.1"
                              /db_xref="GI:14794892"
                              /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                              SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                              VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                              AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                              ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                              KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                              LDDLLLAHSW"
        gene                  5042..33574
                              /gene="amphI"
        CDS                   5042..33574
                              /gene="amphI"
                              /note="polyketide synthase multienzyme polypeptide housing
                              extension modules 9, 10, 11, 12, 13, and 14"
                              /codon_start=1
                              /transl_table=11
                              /product="AmphI"
                              /protein_id="AAK73501.1"
                              /db_xref="GI:14794893"
                              /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                              YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                              GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                              AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                              AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                              GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                              TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                              PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                              EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                              SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                              GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                              PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                              ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
```

DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH

SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH

Query Match   10.6%; Score 45; DB 1; Length 113193;
Best Local Similarity 53.0%; Pred. No. 0.39;
Matches   96; Conservative   0; Mismatches   85; Indels   0; Gaps   0;

Qy   245 gatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaac 304
           ||||||| | | | | ||| || | || | ||| ||| |
Db   92624 GATGGACCTGATGCTGGACGAGTTCCGGGCGGTGGCCGAGACGCTGTCGTTCGCGGCCCC 92683

Qy   305 ggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacgaa 364
           |||| | | || | |||| || | ||| |||| ||| | | |||
Db   92684 GGTGATCCCGGTGGTGTCGAACCTGACGGGTTCGCTGGCCACGGCGGAGGAGCTGTGCTC 92743

Qy   365 gctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgcgagc 424
           || | |||||| | || | || || ||||| | ||||| | ||| | |||| | ||
Db   92744 GCCCGAGTACTGGGTCCGTCACGTCCGTGAGGCGGTCCGTTTCGCCGACGGGGTGAGCGC 92803

Qy   425 c 425
           |
Db   92804 C 92804

RESULT 10
AE004449
LOCUS   AE004449   12829 bp   DNA   BCT   30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 10 of 529 of the complete
            genome.
ACCESSION   AE004449 AE004091
VERSION   AE004449.1 GI:9945928
KEYWORDS   .
SOURCE   Pseudomonas aeruginosa.
  ORGANISM Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1 (bases 1 to 12829)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE   Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2 (bases 1 to 12829)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES   Location/Qualifiers
     source   1. .12829
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene   219. .947
                     /gene="PA0102"
     CDS   219. .947
                     /gene="PA0102"
                     /codon_start=1
                     /transl_table=11
                     /product="probable carbonic anhydrase"
                     /protein_id="AAG03492.1"
                     /db_xref="GI:9945929"
                     /translation="MPDRMRGVKSDSPEQESADDALKRIVDGFQHFRREVFPEQQALF
                     KKLANSQRPRAMFITCADSRIVPELITQSSPGDLFVTRNVGNVVPPYGQMNGGVSTAI
                     EYAVLALGVHHIIVCGHSDCGAMRAVLDPQTLERMPTVKAWLRHAEVARTVVADNCDC
                     GASHDTLGVLTEENVVAQLDHLRTHPSVASRLASGQLFIHGWVYDIESAQIRAYDAKQ
                     GRFLPLDGEHPVPMATPAPRYLSS"
     gene   1158. .2729
                     /gene="PA0103"
     CDS   1158. .2729

```
                        /gene="PA0103"
                        /codon_start=1
                        /transl_table=11
                        /product="probable sulfate transporter"
                        /protein_id="AAG03493.1"
                        /db_xref="GI:9945930"
                        /translation="MHVSKVFPSLRDTLPRDLMASVVVFLVALPLCMGIAIASGMPPA
                        KGLLTGIVGGLVVGFLAGSPLQVSGPAAGLAVLVFELVRTYGVAMLGPILLLAGAIQL
                        LAGRLRLGCWFRVTSPAVVYGMLAGIGILIVLSQLHVMLDLAPKASGLDNLLAFPQAA
                        FAALGSLGMDSGLDAALLGLGTIAVMWGWDKLRPQRLRFLPGALLGVSLATLASLWLA
                        LDVRRVEVPANLGEAIVWLRPADLLALADPSLLLAAVVVAFIASAETLLSAAAVDRLH
                        DGPRSDMDRELSAQGVGNMLCGLLGALPMTGVIVRSSANVNAGARTRASAIFHGLWLL
                        AFVLLLGSLLRQIPVASLAGVLVYTGVKLVDFKALGNLSRYGRMPMLVYAATALAIVF
                        TDLLTGVLVGFALTLLKLVLKAARLKIALRYTGEHEAELRLSGAATFLKVPALSRVLD
                        EVKPGTTLHVPMDNLSYVDHACMELLEDWGRMAPVQGSRLVIEPRALKRRLEGRLRGS
                        VGLGGARNGGAVSPG"
     gene            2866. .3462
                        /gene="PA0104"
     CDS             2866. .3462
                        /gene="PA0104"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="AAG03494.1"
                        /db_xref="GI:9945931"
                        /translation="MVDGFVQPVGEGLADQLVGHQVGYGGVQWDQCLAEVGDVAVVHF
                        FHQAVRQVGFVEQGVEAVVAGEQRRRGEEELLGDLQHRLDPFLDAGFAGHAVGGVEQV
                        RYLFDVGVDEAGEYVFRVLALRLDGAMQVGQAAGYQISQVTVAGFAEVRLLDEFTEGS
                        GVHGVSHTTGMVKKRRNIAGGARRLRRDCHNRFVVMRG"
     gene            3726. .4850
                        /gene="coxB"
                        /note="PA0105"
     CDS             3726. .4850
                        /gene="coxB"
                        /codon_start=1
                        /transl_table=11
                        /product="cytochrome c oxidase, subunit II"
                        /protein_id="AAG03495.1"
                        /db_xref="GI:9945932"
                        /translation="MLRHPRVWMGFLLLSAISQANAAWTVNMAPGATEVSRSVFDLHM
                        TIFWICVVIGVLVFGAMFWSMIVHRRSTGQQPAHFHESTTVEILWTVVPFVILVVMAV
                        PATRTLIHIYDTSEPELDVQVTGYQWKWQYKYLGQDVEYFSNLATPQDQIHNRQAKDE
                        HYLLEVDEPLVLPVGTKVRFLITSSDVIHSWWVPAFAVKRDAIPGFVNEAWTKVDEPG
                        IYRGQCAELCGKDHGFMPIVVDVKPKAEFDQWLAKRKEEAAKVKELTSKEWTKEELVA
                        RGDKVYHTICAACHQAEGQGMPPMFPALKGSKIVTGPKEHHLEVVFNGVPGTAMAAFG
                        KQLNEVDLAAVITYERNAWGNDDGDMVTPKDVVAYKQKQQ"
     gene            4860. .6452
                        /gene="coxA"
                        /note="PA0106"
     CDS             4860. .6452
                        /gene="coxA"
                        /codon_start=1
                        /transl_table=11
                        /product="cytochrome c oxidase, subunit I"
                        /protein_id="AAG03496.1"
                        /db_xref="GI:9945933"
                        /translation="MSAVIDTPDHHAGDHHHGPAKGLMRWVLTTNHKDIGTLYLWFSF
```

```
                          MMFLLGGSMAMVIRAELFQPGLQIVEPAFFNQMTTMHGLIMVFGAVMPAFVGLANWMI
                          PLMIGAPDMALPRMNNFSFWLLPAAFGLLVSTLFMPGGGPNFGWTFYAPLSTTFAPHS
                          VTFFIFAIHLAGISSIMGAINVIATILNLRAPGMTLMKMPLFVWTWLITAFLLIAVMP
                          VLAGVVTMMLMDIHFGTSFFSAAGGGDPVLFQHVFWFFGHPEVYIMILPAFGAVSAII
                          PTFARKPLFGYTSMVYATASIAFLSFVVWAHHMFVVGIPVTGELFFMYATMLIAVPTG
                          VKVFNWVTTMWEGSLTFETPMLFAVAFVILFTIGGFSGLMLAIAPADFQYHDTYFVVA
                          HFHYVLVPGAIFGIFASAYYWLPKWTGHMYDETLGKLHFWMSFIGMNLAFFPMHFVGL
                          AGMPRRIPDYNLQFADFNMVSSIGAFMFGTTQLLFLFIVIKCIRGGKPAPAKPWDGAE
                          GLEWSIPSPAPYHTFSTPPEVK"
          gene            6463. .7017
                          /gene="PA0107"
          CDS             6463. .7017
                          /gene="PA0107"
                          /codon_start=1
                          /transl_table=11
                          /product="conserved hypothetical protein"
                          /protein_id="AAG03497.1"
                          /db_xref="GI:9945934"
                          /translation="MSDAKVDTRRLVGRLLLVTVLMFAFGFALVPLYDVMCRALGING
                          KTAGSAYSGEQQVDVGREVKVQFMTSNNIDMVWEFRSAGDQLVVHPGAVNQMVFYARN
                          PSDKPMTAQAIPSIAPAEAAAYFHKTECFCFTQQVLQPGESIEMPVRFIVDRDLPKDV
                          RHVTLAYTLFDITARKPPVPVAGR"
          gene            7028. .7915
                          /gene="coIII"
                          /note="PA0108"
          CDS             7028. .7915
                          /gene="coIII"
                          /codon_start=1
                          /transl_table=11
                          /product="cytochrome c oxidase, subunit III"
                          /protein_id="AAG03498.1"
                          /db_xref="GI:9945935"
                          /translation="MASHEHYYVPAQSKWPIIASIGLLVTVFGLGTWFNDLTAGHKES
                          HGPWIFFVGGLIIAYMLFGWFGNVIRESRAGLYSAQMDRSFRWGMSWFIFSEVMFFAA
                          FFGALFYVRHFAGPWLGGEGAKGVAHMLWPNFQYSWPLLQTPDPKLFPPPSAVIEPWK
                          LPLINTILLVTSSFTVTFAHHALKKNKRGPLKAWLALTVLLGIAFLILQAEEYVHAYN
                          ELGLTLGAGIYGSTFFMLTGFHGAHVTLGALILGIMLIRILRGHFDAEHHFGFEAASW
                          YWHFVDVVWIGLFIFVYVI"
          gene            complement(7931. .8140)
                          /gene="PA0109"
          CDS             complement(7931. .8140)
                          /gene="PA0109"
                          /codon_start=1
                          /transl_table=11
                          /product="hypothetical protein"
                          /protein_id="AAG03499.1"
                          /db_xref="GI:9945936"
                          /translation="MLKVAIVLLLLATLVSLFSGLFFLVKDQGHGSRVVNSLTVRVVL
                          AAATLVLVAWGFYSGELNSHAPWHF"
          gene            8156. .8950
                          /gene="PA0110"
          CDS             8156. .8950
                          /gene="PA0110"
                          /codon_start=1
                          /transl_table=11
                          /product="hypothetical protein"
                          /protein_id="AAG03500.1"
```

```
                        /db_xref="GI:9945937"
                        /translation="MGGFMQYSQRLGEAGRNDRRSMRGAFRPGLLPTLVVLGLLPVLL
                        WLGTWQLQRADEKRALLASYEARRGAEPVSPGQLEGLRDPAYVRVRLHGRFDERHTLL
                        LDNRLRNGQAGVEVLQPFYDQASGLWLLVNRGWVAWTDRRSPPTLETPDRVLLLDAWT
                        YLPPPGGLHLADAPAGGWPRLVTQLDIPALWQAFGRAGLPWEIRLEPGDASFDTDWPL
                        VSMPPERHTGYAVQWFALATALLALYLYLGVRRAREKNHESRDSDA"
     gene            8925..9503
                        /gene="PA0111"
     CDS             8925..9503
                        /gene="PA0111"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="AAG03501.1"
                        /db_xref="GI:9945938"
                        /translation="MSLAIPMPERPRRARGRLQLLAIIGLVVGPMLLASAMYKWNFWV
                        PQGRSYSGALIGNGQTPADLGVQGRSQGEQWQLLVTAPGTCGEDCQQLVYLARQINIA
                        LGREASRAGHALAASGELPADFANLVRQDYPRLQRYGLDPALYDKAGGGERPLLWVVD
                        PHGNLVLRYDAKANGKKLLKDVQLLLKLSHIG"
     gene            9568..10641
                        /gene="PA0112"
     CDS             9568..10641
                        /gene="PA0112"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="AAG03502.1"
                        /db_xref="GI:9945939"
                        /translation="MSGATRKPGFTVAVLATLLAAVVVLLGAYTRLTHAGLGCPDWPG
                        CYGFVHVPLSEAQLAHAELHFPDAPVEAQKGWNEMIHRYFAGALGLLILGLALHALVR
                        RGRDGQPLKLPLLLLAVVIAQAAFGMWTVTLKLWPQVVTAHLLGGFTTLALLFLLALR
                        LSGRFAARRYPAATRGLAGLALLLVIGQIALGGWVSSNYAAVACIDLPTCHGEWWPRM
```

      Query Match   10.5%;   Score 44.6; DB 1;   Length 12829;
      Best Local Similarity   50.7%;   Pred. No. 0.6;
      Matches 107; Conservative   0; Mismatches 104; Indels   0; Gaps   0;

Qy   185 cgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaatggtg 244
                | | | | | ||| | ||| | ||||||| || | | | | | | | | || | | |
Db   6663 CATGACCTCGAACAATATCGACATGGTCTGGGAGTTCCGCTCCGCCGGCGACCAGTTGGT 6722

Qy   245 gatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaac 304
                | || | || | | | | | | ||| | ||| ||| || | || || | | |
Db   6723 GGTGCATCCCGGCGCGGTGAACCAGATGGTGTTCTACGCGCGCAACCCGAGCGACAAGCC 6782

Qy   305 ggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacgaa 364
                | || | | || | | | | || | | | | | | |||| ||| | ||
Db   6783 GATGACCGCGCAGGCCATCCCGAGCATCGCCCCGGCCGAGGCCGCAGCCTACTTCCACAA 6842

Qy   365 gctcacgtacttgatctggaaccaccatgag 395
                | | | ||| | | ||| || | |
Db   6843 GACCGAATGCTTCTGTTTCACCCAGCAGGTG 6873

RESULT 11
AF015304
LOCUS   AF015304   1766 bp   mRNA   ROD   02-DEC-1997

```
DEFINITION   Rattus norvegicus equilbrative nitrobenzylthioinosine-sensitive
             nucleoside transporter mRNA, complete cds.
ACCESSION    AF015304
VERSION      AF015304.1  GI:2656136
KEYWORDS     .
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 1766)
  AUTHORS    Yao,S.Y.M., Ng,A.M.L., Muzyka,W.R., Griffiths,M., Cass,C.E.,
             Baldwin,S.A. and Young,J.D.
  TITLE      Molecular cloning and functional characterization of
             nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive
             (ei) equilibrative nucleoside transporter proteins (rENT1 and
             rENT2) from rat tissues
  JOURNAL    J. Biol. Chem. 272 (45), 28423-28430 (1997)
  MEDLINE    98019212
REFERENCE    2  (bases 1 to 1766)
  AUTHORS    Yao,S.Y.M., Ng,A.M.L., Muzyka,W.R., Cass,C.E., Baldwin,S.A. and
             Young,J.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (21-JUL-1997) Physiology, University of Alberta, 7-25
             Medical Sciences Building, Edmonton, AB T6G 2H7, Canada
FEATURES             Location/Qualifiers
     source          1..1766
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="jejunum"
     CDS             5..1378
                     /note="rENT1; NBMPR-sensitive; es-type nucleoside
                     transporter"
                     /codon_start=1
                     /product="equilbrative nitrobenzylthioinosine-sensitive
                     nucleoside transporter"
                     /protein_id="AAB88049.1"
                     /db_xref="GI:2656137"
                     /translation="MTTSHQPQDRYKAVWLIFFVLGLGTLLPWNFFITATQYFTSRLN
                     TSQNISLVTNQSCESTEALADPSVSLPARSSLSAIFNNVMTLCAMLPLLIFTCLNSFL
                     HQKVSQSLRILGSLLAILLVFLVTATLVKVQMDALSFFIITMIKIVLINSFGAILQAS
                     LFGLAGVLPANYTAPIMSGQGLAGFFTSVAMICAVASGSKLSESAFGYFITACAVVIL
                     AILCYLALPWMEFYRHYLQLNLAGPAEQETKLDLISEGEEPRGGREESGVPGPNSLPA
                     NRNQSIKAILKSIWVLALSVCFIFTVTIGLFPAVTAEVESSIAGTSPWKNCYFIPVAC
                     FLNFNVFDWLGRSLTAICMWPGQDSRWLPVLVACRVVFIPLLMLCNVKQHHYLPSLFK
                     HDVWFITFMAAFAFSNGYLASLCMCFGPKKVKPAEAETAGNIMSFFLCLGLALGAVLS
                     FLLRALV"
BASE COUNT      338 a    509 c    448 g    471 t
ORIGIN
```

Query Match   10.3%; Score 44; DB 10; Length 1766;
Best Local Similarity   49.2%; Pred. No. 1;
Matches 116; Conservative   0; Mismatches 120; Indels   0; Gaps   0;

Qy   34 atcaggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgtcct 93

```
          | ||   |  ||   |    | |    || ||||| ||  ||||| |  |||    |||| ||
Db   1152 AGCACCACTACCTGCCCTCCCTCTTTAAGCATGATGTCTGGTTCATCACCTTCATGGCCG 1211

Qy     94 ctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgagatcg 153
          |  ||||  ||    ||    || ||||     |  ||||       |||  |    |||  |
Db   1212 CCTTTGCCTTCTCCAATGGCTACCTCGCCAGCCTCTGCATGTGCTTCGGGCCCAAGAAAG 1271

Qy    154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
          | | |   |     |      || |  |    || | ||  |   || | |  | |||| |
Db   1272 TCAAACCGGCTGAGGCAGAGACTGCCGGAAACATCATGTCCTTCTTTCTGTGTCTGGGCC 1331

Qy    214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactg 269
          ||| |||   |||||    ||| |    | ||  | ||  | |||  | || |  |||
Db   1332 TGGCTCTGGGAGCTGTGTTGTCCTTCTTGTTAAGGGCACTTGTGTGAGCGACCCTG 1387


RESULT   12
CELF56C9
LOCUS       CELF56C9     35028 bp     DNA              INV        14-MAR-2001
DEFINITION  Caenorhabditis elegans cosmid F56C9, complete sequence.
ACCESSION   U00063
VERSION     U00063.1  GI:488186
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 35028)
  AUTHORS   The C. elegans Genome Sequencing Consortium, Washington University
            Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
            Hinxton, U.K.,C.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
REFERENCE   2  (bases 1 to 35028)
  AUTHORS   Du,Z.
  TITLE     The sequence of C. elegans cosmid F56C9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 35028)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1994)
REFERENCE   4  (bases 1 to 35028)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
```

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
    source           1. .35028
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="CELF56C9"
    gene             complement(1504. .2040)
                     /gene="F56C9.5"
    CDS              complement(join(1504. .1680,1731. .1927,1974. .2040))
                     /gene="F56C9.5"
                     /note="similar to Acyl-CoA-Binding protein; most similar
                     to ACBP region of B. taurus endozepine-related protein;
                     contains similarity to Pfam domain PF00887 (ACBP),
                     Score=184.8, E-value=4.6e-52, N=1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAK18960.1"
                     /db_xref="GI:13324982"
                     /translation="MGKSLDEQFEAAVWIINALPKNGPIKTSINDQLQMYSLYKQATS
                     GKCDTIQPYFFQIEQRMKWNAWNQLGNMDEAEAKAQYVEKMLKLCNQAEAEHNLMEFL
                     SDPTIADLLPKQNQLREHFATLGRTTVKGFEGETVEINGVSISF"
    gene             complement(2294. .3866)
                     /gene="F56C9.6"
    CDS              complement(join(2294. .2336,2384. .2477,2548. .2710,
                     2753. .2950,2997. .3089,3134. .3261,3313. .3388,3508.
.3572,
                     3622. .3688,3762. .3809,3852. .3866))
                     /gene="F56C9.6"
                     /note="coded for by C. elegans cDNA CEESY75F"
                     /codon_start=1
                     /product="Hypothetical protein F56C9.6"
                     /protein_id="AAK18961.1"
                     /db_xref="GI:13324983"
                     /translation="MERVNERREQNNPNGCCLRDEDFSQFNSEAFLRELAADLNEDDT
                     NDLSSSLFATSRIPEEHIRSTGLVERAEHYNKSVDQRTMTDARIAFEEEQLKNGKSPN
                     AGTSGMENLADSGTHVPRRGRGDYFGKLRSFENGVSFPSRPPLTSEHSSSGDSYFNNS
                     HKTTPNYRRFANSNDSSRDNSQMEYKAENDASHTSQSSNRFGFNSQINRTDIHPPAAR
                     HTFNPAAAYNGKITPDFRFNYIPNAAVPAPSVVPVIATHPGVAPPSIVPSPIRIGQRY
                     PKRPDNMPKPSSEPKHLNHNYYQIELYGATQEDRIAQRIEKTVRQTEAPVRRF"
```

```
gene            4478. .6064
                /gene="F56C9.3"
CDS             join(4478. .4516,4572. .4622,4707. .5280,5325. .5427,
                5481. .5859,5903. .6064)
                /gene="F56C9.3"
                /note="weakly similar to R. rickettsii protein P34; coded
                for by C. elegans cDNA cm13h9; coded for by C. elegans
                cDNA yk195g2.3"
                /codon_start=1
                /protein_id="AAK18959.1"
                /db_xref="GI:13324981"
                /translation="MSDSSQTEHKNPEREQFKFKFHGQYHGQYHAQAAKKQLTEYYKK
                QNEILDHFKQDSEQIEATRRTKIRHQSLKSNESEFSEVNEHDHLSSLKASTVSIHSKD
                SLMVRHEEAQNEEIKLTKAAARLAHITLFVNLVLMLAKIFASYLSGSMSIISSMVDSV
                VDLTSGAVLSISSRMIRKRDPYQYPRGRTRVEPLSLILISVIMGMASVQLIISSVRRI
                HDAAVYGIKDPINVSWPTIAIMGSTIAVKLTLFIICQKYKSNSSIKVLSLDHRNDCIS
                NSMALACAWLAFYYTVKDGDEKSGAVVFEKQFDLYYLDPAGAILVSVYILYTWIRTGY
                AHFVMLSGKSAHPELINRIVHQCIEHDPRITHIDTVYVYHYGTKFLVEVHIVLDQNMS
                LKVTHDIAESLQTGIESLPEIERAFVHCDYEFEHHPHDEHKAV"
gene            complement(6359. .8125)
                /gene="F56C9.7"
CDS             complement(join(6359. .6606,7239. .7396,7451. .7572,
                7940. .8125))
                /gene="F56C9.7"
                /note="coded for by C. elegans cDNA yk367h2.3; coded for
                by C. elegans cDNA yk367h2.5"
                /codon_start=1
                /product="Hypothetical protein F56C9.7"
                /protein_id="AAK18962.1"
                /db_xref="GI:13324984"
                /translation="MLLRFIIGLSSVLLVTGLFNPLLPTLGDYLLTDQQVQKYFSVAG
                PSNLAIGTCNSVPFASAQAGFASSVGLAPTTTWREANILTNATIGMIDQGMDQLAAVC
                QARQQFVQTLGAAYDTCTDRFYLISLGNTDWWNVMQYTHLMKHLEFICSTGFDVYQSN
                IDCIRKGETTDGNQYRACFYKFNATVNANPNNFCGATETFIGCIKDFFDTECNLYVGW
                MQCELERIGFAYDCYGLSC"
gene            complement(9629. .11911)
                /gene="F56C9.8"
CDS             complement(join(9629. .9899,10182. .10238,10298. .10374,
                11452. .11538,11855. .11911))
                /gene="F56C9.8"
                /note="coded for by C. elegans cDNA CEESJ95F"
                /codon_start=1
                /product="Hypothetical protein F56C9.8"
                /protein_id="AAK18963.1"
                /db_xref="GI:13324985"
                /translation="MGYVPAYTFDEPFAGNGPQIAVVFAIVIGGVLLLAAFIYVIYAA
                VVRSMRSDDKQRLHSGRNSAQWNQPQQQYREQQPASDFPYNPAPTQNYDYNAPIRTPV
                NPTSFTPVPSVTQYSTQPQQYSNVPLVTPTTQQYIQNQSIPQYAPDVIYTQGPTGYIQ
                QQPIEYNVQRQSPGANFVQSSV"
gene            complement(18269. .19278)
                /gene="F56C9.9"
CDS             complement(join(18269. .18451,19045. .19278))
                /gene="F56C9.9"
                /codon_start=1
                /evidence=not_experimental
                /product="Hypothetical protein F56C9.9"
                /protein_id="AAK18964.1"
```

```
                        /db_xref="GI:13324986"
                        /translation="MEKKRVSWRLTTCHVEFLRYSTPDASLGPSRDWVFHVPRWNSLF
                        FNSDLVVFRSWPPAVLKGEDRGPSELDCRVVDPSSLSEQARNTINRLILMEPNTQKRG
                        ISDETVSNASPPEKSPRLVNSPMPAMNQLLSPILPN"
         gene           19444. .21974
                        /gene="F56C9.2"
         CDS            join(19444. .20161,20374. .21974)
                        /gene="F56C9.2"
                        /note="similar to reverse transcriptase; contains
                        similarity to Pfam domain PF00078 (rvt), Score=61.1,
                        E-value=7.8e-15, N=1"
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAK18958.1"
                        /db_xref="GI:13324980"
                        /translation="MVLNVYAPVCKSRSENDAKRFFERLRAEYFQLRKSFKGPIVLGG
                        DFNAATSCSTNDELAPWICGNVFGNSNNHGDFFFNFLASTRLFQLNSRFPKRLAKRWT
                        FAGKKAVGRTEIDFFISSRIDLVKDVSTFSNLHNLSDHRLIRSRWAISVKSERDHAFK
                        SRRLPSTGKERDCTLYEDAIKDLSNHATFGSYDSFVKTLRQGLVPLPAHKPQKFSQRT
                        NMILQERRSVLESSAPDTAKLRTIRGSHELSTSTTNPNPPSLDPIPAILKSEVRLEIR
                        KLKTKSAPGLDNVDAAMLKNGGDTVVDSVTALFNNILQHNKVPDLWKISDVKLIPKKA
                        KATKIKDFRPISLLPILSKMFSSILTRRLTPTLESYLDESQNGFRKGRCCADNIQSLT
                        MLIEKCNEFQLPLLLLFIDYQTAFDKIGHSAVVSSLEKAGADPAMRKMIQEMMDGGQA
                        EITVHDKKLKVNLCTGVRQGDSASPALFSAALQAILTDCDNEFAGVGINVEGRHIRRL
                        EFADDVVLICSTPGEVQERLEILDRISSNYGLKINQSKTVLLKNKFCRSQDVLFNGSP
                        IIPVPGCRYLGRWIDISGSIDEEISRRIRAGWGALVGIKEVLRIMPNKERIILFKQNV
                        LPALLYASETWTCNAGSTLRLKRTVSGLIDAAEIRGWNFNLDRYLLAKQSRFTGHILR
                        RDPNRWTKICTEWDPSHNKNWKRAVGGQKKRWAKDIDEEYAKFHHNSAMSGQVVVGRR
                        RLGMLTPKVPWLSIARTDREKWKEFVRSCLAT"
         gene           complement(23375. .28267)
                        /gene="F56C9.10"
         CDS            complement(join(23375. .23701,23819. .23949,24004. .25339,
                        25386. .25481,25563. .25973,26090. .26572,26758. .26940,
                        28148. .28267))
                        /gene="F56C9.10"
                        /note="coded for by C. elegans cDNA yk7g2.3; coded for by
                        C. elegans cDNA yk7g2.5; coded for by C. elegans cDNA
                        yk27a4.3; coded for by C. elegans cDNA yk27a4.5; coded for
                        by C. elegans cDNA yk45d1.5; coded for by C. elegans cDNA
                        yk152g1.5; coded for by C. elegans cDNA yk154d4.5; coded
                        for by C. elegans cDNA yk154d4.3; coded for by C. elegans
                        cDNA yk193d3.3; coded for by C. elegans cDNA yk193d3.5;
                        coded for by C. elegans cDNA yk226f10.3; coded for by C.
                        elegans cDNA yk226f10.5; coded for by C. elegans cDNA
                        yk243h12.3; coded for by C. elegans cDNA yk243h12.5; coded
                        for by C. elegans cDNA yk333a9.3; coded for by C. elegans
                        cDNA yk333a9.5; coded for by C. elegans cDNA yk347e3.5;
                        coded for by C. elegans cDNA yk396c7.3; coded for by C.
```

```
    Query Match            10.3%;   Score 44;   DB 3;   Length 35028;
    Best Local Similarity   53.5%;   Pred. No. 0.76;
    Matches    92; Conservative   0;  Mismatches  80;  Indels    0;  Gaps    0;


Qy    247 tggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacgg 306
          |||| |||| ||||||||||||  |   |  | |||||   ||   |||||||  ||||| |
Db   5651 TGGATCCTGCTGGCGCCATACTTGTATCTGTATACATCCTCTACACGTGGATCCGAACCG 5710
```

```
Qy    307 tgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagc 366
              | | |  | | | ||  || |   || ||||| || |    ||||||   | |      |
Db   5711 GATACGCGCATTTCGTCATGCTCAGTGGAAAGTCAGCTCATCCAGAGTTGATCAATCGGA 5770

Qy    367 tcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggt 418
              |     |    |      | ||  ||     |||     || || |||||| ||
Db   5771 TTGTTCATCAGTGTATCGAGCATGATCCACGGATTACACATATTGACACCGT 5822
```

RESULT 13
AE005086/c
LOCUS   AE005086   11548 bp   DNA   BCT   12-FEB-2001
DEFINITION   Halobacterium sp. NRC-1 section 117 of 170 of the complete genome.
ACCESSION   AE005086 AE004437
VERSION   AE005086.1 GI:10581303
KEYWORDS   .
SOURCE   Halobacterium sp. NRC-1.
  ORGANISM   Halobacterium sp. NRC-1
             Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
             Halobacterium.
REFERENCE   1 (bases 1 to 11548)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
             Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
             Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
             Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
             Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
             Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
             Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
             Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
             DasSarma,S.
    TITLE   From the cover: genome sequence of halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
   PUBMED   11016950
REFERENCE   2 (bases 1 to 11548)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
             Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
             Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
             Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
             Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
             Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
             Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
             Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
             DasSarma,S.
    TITLE   Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
             Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES   Location/Qualifiers
     source   1. .11548
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene   complement(287. .1528)
                     /gene="VNG1857C"
     CDS   complement(287. .1528)
                     /gene="VNG1857C"
                     /note="conserved hypothetical protein"
                     /codon_start=1

```
                        /transl_table=11
                        /product="Vng1857c"
                        /protein_id="AAG20061.1"
                        /db_xref="GI:10581304"
                        /translation="MHSTTRREWLGAIGATAATGLAGCAGVGGAGQPVTVGSLLPLSG
                        PGSLGALAADHQRAIDTAVEHANRGGGINGRDVVHVSKDTEADPSVAADRYATLAADE
                        SPLAIVGPVLSGVTTALTEQAAADAQLLVSPSTTAPAIATAGRSDGQKFVARTCPNDS
                        QQAAVMAKIVDDDMYAAADTATILYVDNAFGAALADVLADRLGADLLASVPYQGGTDT
                        PGGPVDDALAPDPDAVAFIGSPGSSSGVIDELVGREYGGEIALSSALASASSPPSWNG
                        AYTATVNSASTVGTKRLRRALSDATPLQPYTENAYDAAALALLAASYSGDPTPRAVAG
                        ALQSVSGGVGHSITVGDFGRATDLIDAGRELNYNGATGNVDLTAALEPVTGYLIQQLT
                        DAGIETRELLKSGYFTDGGDA"
gene                    1626. .2264
                        /gene="deoC"
                        /note="VNG1859G"
CDS                     1626. .2264
                        /gene="deoC"
                        /note="DeoC"
                        /codon_start=1
                        /transl_table=11
                        /product="deoxyribose-phosphate aldolase"
                        /protein_id="AAG20062.1"
                        /db_xref="GI:10581305"
                        /translation="MDRETLAARIDHTVLGPTTTRADVLSVVDDAEAHGMNVCIPPCY
                        VADARDHASADRTIATVIGFPHGTQATSVKVAAAEHAHADGADELDLVIPIGRLKGGD
                        HEAVTAEIAAVNDATPLPVKVIIETPVLTDAEKHAACEAAADADAAMVKTATGFTDGG
                        ATVPDVSLMSEYLPVKASGGVGTYADAAAMFDAGAVRIGASSGVDIVASFAE"
gene                    2323. .3621
                        /gene="VNG1861C"
CDS                     2323. .3621
                        /gene="VNG1861C"
                        /note="conserved hypothetical protein"
                        /codon_start=1
                        /transl_table=11
                        /product="Vng1861c"
                        /protein_id="AAG20063.1"
                        /db_xref="GI:10581306"
                        /translation="MARYHIETYGCTSNRGESRDIERRLRDAGHHKVETAADADVAIL
                        NTCTVVEKTERNMLRRAKELADETADLIVTGCMALAQGEAFADADVDAQVLHWDDVPE
                        AVTNGECPTTTPDAEPILDGVVGILPIARGCMSNCSYCITKQATGRVDSPPVEENVEK
                        ARALVHAGAKEIRITGQDTGVYGWDTGERKLPELLERIATEIEGEFRVRVGMANPGGV
                        HGIREELAAVFAEHDEIYNFLHAPVQSGSDDVLADMRRQHEVSQYRDIVETFNDTLGE
                        WTLSTDFIVGFPTEDDDDHEASMDLLRETRPEKINVTRFSKRPGTDAAELKGLGGQTK
                        KDRSKAMTELKMDVVGEAHESMVGTRRDVLVVEEGTGDSVKCYDGAYRQVIVQNATDH
                        GLEPGDFATVEVTSHQTVYAFAEPVDAAAVDDGPAETTAD"
gene                    complement(3640. .4539)
                        /gene="cef"
                        /note="VNG1862G"
CDS                     complement(3640. .4539)
                        /gene="cef"
                        /note="Cef"
                        /codon_start=1
                        /transl_table=11
                        /product="cation efflux system protein (zinc/cadmium)"
                        /protein_id="AAG20064.1"
                        /db_xref="GI:10581307"
                        /translation="MERKRAVRRVGALVLAANLALVAAKGAAWWATGSLAVGSEAINS
```

```
                           LADVAYSLVVLGGLYLTTQPPDFKHPHGHERIEPFVSLVVALGVLAAGGAVLWQATTT
                           VAAGDYGPTPGLPAVGVLVGTAVAKYALYRYVLGVAADHRSPALRATALDNRNDILTA
                           SAALVGVLGSATGYPVLDPLAAFVVAAGILHTGYEIVRDNVNYLVGAAPPADLREQIL
                           GRALDNPDVEGAHDVVAHYVGPEIDVSLHVEVEGEMTLHEAHDIETDLILDLESIPEV
                           DDVFVHVDPKELGEWKDADTAPE"
     gene                  complement(4580. .5125)
                           /gene="hit2"
                           /note="VNG1864G"
     CDS                   complement(4580. .5125)
                           /gene="hit2"
                           /note="Hit2"
                           /codon_start=1
                           /transl_table=11
                           /product="histidine triad protein"
                           /protein_id="AAG20065.1"
                           /db_xref="GI:10581308"
                           /translation="MEQVFAPWRIEWVERDDTTDDDVDCVFCAFPGREHARQHLVVAR
                           TDHAAVMLNNYPYNPGHCMVIPDVHTGDYGDLDADTLLDHARLKQATLDALDAALGPD
                           AFNTGLNLGGGAAGGSIGDHLHTHVVPRWNGDTNFMPVISDTKVIVEALDDTYDRLHD
                           AFLARPDATEAPTGAVFVDFN"
     gene                  5246. .5446
                           /gene="VNG1865H"
     CDS                   5246. .5446
                           /gene="VNG1865H"
                           /note="hypothetical protein"
                           /codon_start=1
                           /transl_table=11
                           /product="Vng1865h"
                           /protein_id="AAG20066.1"
                           /db_xref="GI:10581309"
                           /translation="MASKTPGFEGVTEYCERCGQTTTHQVAVELRTENTNTENAAFSR
                           EPYRVATCCECDAEHAQRMNNA"
     gene                  complement(5484. .6374)
                           /gene="map"
                           /note="VNG1866G"
     CDS                   complement(5484. .6374)
                           /gene="map"
                           /note="Map"
                           /codon_start=1
                           /transl_table=11
                           /product="methionyl aminopeptidase"
                           /protein_id="AAG20067.1"
                           /db_xref="GI:10581310"
                           /translation="MTDSVTVGSDAYEQYVEAGDILTTVLSEAADRVTVGATHLEVAS
                           FAEERIRELGGEPAFPVNISVNEEASHAAPGADDDTEFGEDMVCLDVGVHVDGHIADA
                           ATTVDLSGTPELVEAAEESLAAAIDMVEPGVQTGALGAEIQDVVEAYGYNPVVNLTGH
                           GMDVFDAHTGPTVPNRGVDSGAELAVGDVVAIEPFVTTGTGKVTEGAATEIYEVVSSG
                           TVRDRRARQLLDDLEQFDGLPFAARWLDGARAEMSLTRLERADIVRSYPVLKEADGEL
                           VGQDEHTLIVTEDGCEVVTA"
     gene                  complement(6423. .7238)
                           /gene="potC"
                           /note="VNG1867G"
     CDS                   complement(6423. .7238)
                           /gene="potC"
                           /note="PotC"
                           /codon_start=1
                           /transl_table=11
```

```
                              /product="spermidine/putrescine ABC transporter permease"
                              /protein_id="AAG20068.1"
                              /db_xref="GI:10581311"
                              /translation="MASPSTRSRLLGHALSAWTVAVLAVLWLPLVVIIVLSVAENAAT
                              ILPFTGVTLAHYQATLQDGALLGSVANSATIATLASVLATAVGVPASVALVRYDVPLS
                              NAFRVAVVLPMVVPGVVLGIGVLISIRTLPGITPGFVPTVLTHAVYGLPFVVLLVSAR
                              LAAVDDTLADAARDLGASPLVAFRDVTLPAIAPAVASGFLFAWVRSFEEFVRAYFVSG
                              TTDVLTTEMYALLAYGTAPKLNVIATLVLFVLAVVLAVAMTAGDVVSAVTAGE"
          gene                complement(7238..8131)
                              /gene="potB"
                              /note="VNG1868G"
          CDS                 complement(7238..8131)
                              /gene="potB"
                              /note="PotB"
                              /codon_start=1
                              /transl_table=11
                              /product="spermidine/putrescine ABC transporter permease"
                              /protein_id="AAG20069.1"
                              /db_xref="GI:10581312"
                              /translation="MGDTDSVLGRLVGSPRARLAALLAPSGGLLVALLFAPLSFMVAV
                              SFARVSDASRIIWHPTAANYTALVDATPFWSTPFVTSLLLSVGIAAATTVVCLVAAYP
                              VAYALARRDRGRRVVFFLVLLPFFTMYLVRVYSWYLLFGDGGVLNDIATVLGVGPVDA
                              FGFGVPAIVVGLAHAQFPYMLLTLYAGIEAVDFDVLEAARDLGASRRAVFRDVLLPLT
                              LPNVVAGSLFVFVPAFGSFVAPRFLSGSTVLLVGQLIAGRIDSFNIASASAAATVVVV
                              LIAAAFGVAARYTDASAGGDH"
          gene                complement(8133..9311)
                              /gene="VNG1869C"
          CDS                 complement(8133..9311)
                              /gene="VNG1869C"
                              /note="conserved hypothetical protein"
                              /codon_start=1
                              /transl_table=11
                              /product="Vng1869c"
                              /protein_id="AAG20070.1"
                              /db_xref="GI:10581313"
                              /translation="MSPRQPDASDADAPADTTSDAGHPHSTVGRRSFLAATGAAASAT
                              TLAGCLGGTTGTPTINVLTWEHYARDDVAAAVEDAVDATVNVTKSISSEKMFAGWQAG
                              KDQQFDIAIPNNNYVPKFAAAGLAAPVPVSDLDSYPAIYDFFKNAMDTQLAVDGTPYG
                              VPIRFGYYGYGYDSRAVPDHDPSWSVLFDGIDGVDLSSAVALYDNHFKAMSAAALHLG
                              FRDAFDGDRITLSESQLGAVTDALIDQKQSLLSGYIAGDASFVKGITKGDFHVGHTGR
```

```
          Query Match                9.2%;   Score 39;   DB 1;   Length 11548;
          Best Local Similarity    48.8%;   Pred. No. 14;
          Matches   105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;


Qy     154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
           || | ||| ||||| ||| |   |  | |||| | |||| |  | | | ||||||
Db    4103 TGCGCGCCACCGCCCTCGACAACCGCAACGACATCCTCACCGCCAGCGCGGCGCTGGTCG 4044


Qy     214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatcg 273
           |  || | | | |  | ||    ||| |||||||| ||| || | | ||
Db    4043 GCGTCCTCGGCTCCGCCACCGGCTACCCCGTCTTGGACCCGCTTGCGGCGTTCGTCGTCG 3984


Qy     274 cgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtaggcacactgatag 333
           |    || || ·||  | |    | || | ||| ||||| | | |||| |
Db    3983 CCGCCGGCATCCTCCACACCGGCTACGAGATCGTCCGGGACAACGTCAACTACCTGGTCG 3924
```

```
Qy    334 gcaagtcggcgccggcagagtacctgacgaagctc 368
          ||    || |||| || ||   ||   | || ||
Db   3923 GCGCTGCGCCGCCCGCCGACCTCCGCGAGCAGATC 3889
```

RESULT 14
RSCHECTOR
LOCUS   RSCHECTOR   14713 bp   DNA   BCT   13-OCT-2000
DEFINITION Rhodobacter sphaeroides aam1 gene (partial), ORF7, cheY5 gene, mcpB
            gene, tlpS gene, mcpA gene, cheD gene, cheY1 gene, cheA1 gene,
            cheW1 gene, cheR1 gene, cheY2 gene, ORF2 and ORF3 (partial).
ACCESSION   X80205 X86707
VERSION   X80205.3 GI:7573209
KEYWORDS   aam1 gene; alpha amylase; cheA1 gene; cheD gene; chemotaxis
            histidine protein kinase; chemotaxis response regulator; chemotaxis
            scaffold protein; cheR1 gene; cheW1 gene; cheY1 gene; cheY2 gene;
            cheY5 gene; mcpA gene; mcpB gene; methyl accepting chemotaxis
            protein; ORF2; ORF3; ORF7; tlpS gene; transducer like protein.
SOURCE   Rhodobacter sphaeroides.
  ORGANISM Rhodobacter sphaeroides
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodobacter.
REFERENCE   1 (bases 1 to 14713)
  AUTHORS   Ward,M.J., Bell,A.W., Hamblin,P.A., Packer,H.L. and Armitage,J.P.
  TITLE   Identification of a chemotaxis operon with two cheY genes in
            Rhodobacter sphaeroides
  JOURNAL   Mol. Microbiol. 17 (2), 357-366 (1995)
  MEDLINE   96079285
REFERENCE   2 (bases 1 to 14713)
  AUTHORS   Ward,M.J., Harrison,D.M., Ebner,M.J. and Armitage,J.P.
  TITLE   Identification of a methyl-accepting chemotaxis protein in
            Rhodobacter sphaeroides
  JOURNAL   Mol. Microbiol. 18 (1), 115-121 (1995)
  MEDLINE   96154945
REFERENCE   3 (bases 1 to 14713)
  AUTHORS   Shah,D.S., Porter,S.L., Martin,A.C., Hamblin,P.A. and Armitage,J.P.
  TITLE   Fine tuning bacterial chemotaxis: analysis of rhodobacter
            sphaeroides behaviour under aerobic and anaerobic conditions by
            mutation of the major chemotaxis operons and cheY genes
  JOURNAL   EMBO J. 19 (17), 4601-4613 (2000)
  MEDLINE   20428429
REFERENCE   4 (bases 1 to 14713)
  AUTHORS   Ward,M.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-AUG-1994) M.J. Ward, Oxford University, Microbiology
            Unit, Biochemistry Dept, South Parks Rd, Oxford, UK
  REMARK   Revised by [5]
REFERENCE   5 (bases 1 to 14713)
  AUTHORS   Porter,S.L.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-APR-2000) Porter S.L., Department of Biochemistry,
            Microbiology Unit, University of Oxford, South Parks Road, Oxford,
            OX1 3QU, UNITED KINGDOM
COMMENT   On Apr 14, 2000 this sequence version replaced gi:7532750.
FEATURES   Location/Qualifiers
     source   1. .14713
                     /organism="Rhodobacter sphaeroides"
```

```
                    /strain="WS8N"
                    /db_xref="taxon:1063"
gene                complement(1. .2034)
                    /gene="aam1"
CDS                 complement(<1. .2034)
                    /gene="aam1"
                    /codon_start=1
                    /transl_table=11
                    /product="putative alpha amylase"
                    /protein_id="CAB87126.1"
                    /db_xref="GI:7532751"
                    /translation="MPKMRMALAAPSAGRGNQRLCDRLSSHTRPRWGSGVTDRNGESG
                    RMRPARALRAPKGPDSEANRQGVAAATGAQDRRDGGEEALMRLADARVAIEGVNLEID
                    GGRFAAKVVAGWEVAVEADIFCDGHDSIDAAVLHRQRGTDDWTEVRMEFLVNDRWQAR
                    VTFAENAFHELTFLAWRDLYTTWRKEVAKKLAAGQKIDLELEEGRRLLQSVETAGAED
                    RALVDRILGEDGADQEAGARFARMSSPEAVAAMKRCAPRTNLTCYKILPIFADREAAA
                    FSAWYEMMPRSQSGDPERHGTFDDVIRKLPYVRDLGFDVLYFTPIHPIGRVNRKGRNN
                    SLTPGPDDPGSPYAIGSEEGGHDAIHPELGDFESFGRLVEAAHAHGLEVALDFAIQCA
                    PDHPWIREHPEWFDWRPDGTIKFAENPPKKYEDIVNVHFYRGALPELWYALRDVVLFW
                    VEKGVKIFRVDNPHTKPFPFWEWMIGEVQSQHPDVIFLAEAFTRPKVMKRLGKVGYGQ
                    SYSYFTWRNTKAELIDYLTELTTEECRHYMRPNFFANTPDINPVYLQHSGRAGFRVRL
                    ALAATLGGNYGLYNGYEICEATPVPGKEEYFNSEKYQLRAWDFDQPGHIQDDIRLMNH
                    IRRTHPAMRDFTRLRFYDAHNDSVLAYGKSTEDKQDFLLFHVNLDPHAAQTFEF"
CDS                 2061. .2444
                    /note="ORF7"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="CAB87127.1"
                    /db_xref="GI:7532752"
                    /translation="MGSGPRTAGRKQSAEAPAMSAPPVSFSLPDRAGLPEAGPLAADL
                    ARAFAGPGPVRLDTAPAQEVGLAVLQLLVAAHRQAASAGVQFEIAVPAGSPMETAMKV
                    HGLADAGLVGADGLWTGLPVGVAQG"
gene                2441. .2809
                    /gene="cheY5"
CDS                 2441. .2809
                    /gene="cheY5"
                    /citation=[3]
                    /codon_start=1
                    /transl_table=11
                    /product="chemotaxis response regulator"
                    /protein_id="CAB87128.1"
                    /db_xref="GI:7532753"
                    /translation="MSKTILAVDDSPSVRQMVRLTLVGAGYTVVEAVDGQDALEKATA
                    QRFDAILTDQNMPRLDGIGFIRKFRTLPEGKGVPIVFLSTESQDTLKAQAKEAGAIGW
                    MIKPFDQAQLLAVVKKVAGA"
gene                2898. .4580
                    /gene="mcpB"
CDS                 2898. .4580
                    /gene="mcpB"
                    /codon_start=1
                    /transl_table=11
                    /product="methyl accepting chemotaxis protein"
                    /protein_id="CAB87129.1"
                    /db_xref="GI:7532754"
                    /translation="MRLSIKLKLAGVFLAVLLVSGGGQMVALRDLDQIRASLDDIVHT
                    KVKQVEMTYQLIENRLKTQREIRNYLLSRTKEERRAIDDRLATASAGSEQAFAALEAS
```

```
                    ADAETRARLAEVQEAKERLARIDEKAIEMARMGLGYEGFTIVVTQGREQWLAMETRLS
                    ALLAHHTQQLTDASAEAQRQQEISRLTVLGAFLANILLVAAAGSWIVVTLSTGLKRAL
                    RLSERVAAGDLSQTEPQSQRDEIGDLIASLNGMVTKLRTVVNDVARSTRTVAAGADEM
                    SSTAVKLSQGAAEQASATLQASSSMEEMTANIKQSAQNAADTDTRARQSALAAHESGT
                    TMVEAVEAVRTISQKIGIVQEIARQTDLLALNAAVEAARAGEHGRGFAVVAAEVRKLA
                    ERSRAAATEISVLSAATVEAAQTAGGRLSQLVPDIEETARLVLEISTSAQEQAAGVAQ
                    VNTAIQQLDQVTQSNSTASEQLSATAGQLAGQAEQLRTAIGFFTTDRGPEVVPSHESL
                    PTAQLPTRSELPRRSTKSRPVTQSPGGFQFDLDGNGDDLDADFRRHATDHAA"
     gene           4684. .6471
                    /gene="tlpS"
     CDS            4684. .6471
                    /gene="tlpS"
                    /codon_start=1
                    /transl_table=11
                    /product="transducer like protein"
                    /protein_id="CAB87130.2"
                    /db_xref="GI:7573210"
                    /translation="MALSTPLPGEATATAHQPGPVAPPGEGTFDRANGRRRVSTPAMR
                    KAASRVSGLASATEKAFLRAGASLERAISRFDRLSAPLAQLAAVADAGEFAQASADAA
                    GLEERAALFAANSGPLLERIETLSAAAETLGTDLSVMRQVIRTMSIVALNARVTVATL
                    AGQNSSLEVFTTAATAQVMEAGDTIGQITEAVEHMARRLHLASAEAGNLSHLLRRQLG
                    PALAGLRLDMDAFEKDLTRTTADGSILMTHGDEFRRAVTTAVLSLQIGDTTRQRLEHV
                    ADMMDGTAAQETADGALARITLALAAAHLRDAHERHAAAIVTARSALRDSGQATAEIG
                    RISGRVGTSPRHNLKQHLQQLQAILDECRTAQTRLVSVARNLSSGLAELLPILERMSG
                    VEERMSMIGLNAVIACVQLGDEALALREISFQLRELASTSAERLGSITRSLSAMSDEA
                    VITAVELEGPFKQDLHDLTGAGDRVFSLLSGIEAGILQTGATVERERRLAERDVATGI
                    AALDGHSAAFADLLTMAPALERWANRLDGDLNATATGDTLERIRAGYTMAAERLVHDL
                    LLRDLGVNPAEEVSETTAPPADDILDVLF"
     gene           6578. .8956
                    /gene="mcpA"
     CDS            6578. .8956
                    /gene="mcpA"
                    /citation=[2]
                    /codon_start=1
                    /transl_table=11
                    /product="methyl accepting chemotaxis protein"
                    /protein_id="CAB87131.1"
                    /db_xref="GI:7532756"
                    /translation="MNRTDPSRLWRFKQPLVLLGTPLPMLLGIATALWLNHEGALRAT
                    EQQTQAFSRLIAAEGRAALAFRDGHRLSQLFGTAAATYPGETFSALAMDVEGRVIASM
                    PADLAGAENLQAQALSAMTAGAPVKAKGGTSIAIPVTREGDDSLAGVFAVSLPALNGS
                    YALLMPATALVAGLIISMGIAGHLWRRRKETERLLIETTRRIKNNQRPDATAMTRIEL
                    SIPTLANEIDALSAALQGEREQFEAAHSRAMALDALPAPWILVASDGRVLMMNHPARQ
                    VVAGLPEPLLEGQPVASLHHDLARAWPDRTGRTMSLAIGGRQYQVTRAPVGATGAEVL
                    SFTDRTEETQLDLLLQGVIREAIAATYDIRGQMIKASDGFAKLFGSSGASLRSLLEAT
                    PDSAELLAAVEREGEGRSFLSREGYGSDQVQVGISILRRPGGGHLVLVTEIRHQYAQA
                    PTTAAADQPPPAPERLISALRQLAQGDLGSRLDHPLPEPFEALRPDFNSALQGLASLV
                    EDVISAAESIRNEARDISSAAQSLAQRTESTAATLEETAAALDGLTVSVRSAADGAAE
                    ADRVVADARANAEESGHVVVETVAAMDMIAASSDKITSIVKVIDDIAFQTNLLALNAG
                    VEAARAGDAGRGFAVVASEVRALAQRSSEAAREITDLILKSGNQVRRGVDLVGKTGDA
                    LKQIVSSVSEISTLVSDIAVSSRQQSVSLAEINCAVNNLDQSTQQNAARLEEATAASE
                    SLTTSANALFETVQQFHLDAPPKRNRPTPPLTAATPHNSRALARAEPGWEDF"
     gene           9014. .9613
                    /gene="cheD"
     CDS            9014. .9613
                    /gene="cheD"
                    /function="putative role in chemotaxis"
```

```
                             /codon_start=1
                             /transl_table=11
                             /product="CheD protein"
                             /protein_id="CAB87132.1"
                             /db_xref="GI:7532757"
                             /translation="MTRCDDRPSASQISITHVTQGSCVASSSPNEVYATILGSCICTC
                     MCDPVAGVGGMNHFLLPSADVEDAQHLRYGSHAMELLINALLKLGAARQRIEAKIFGG
                     AMMTPQLGAIGQANAAFARRYLRDEGIRCTAHSLGGNRARRIRFWPKTGRVQQMFLGS
                     EDVVPNEQPQFRLQGGAGDVTFFDRHNNAEMPDPIKEPR"
     gene                    9885. .10253
                             /gene="cheY1"
     CDS                     9885. .10253
                             /gene="cheY1"
                             /citation=[1]
                             /citation=[3]
                             /codon_start=1
                             /transl_table=11
                             /product="chemotaxis response regulator"
                             /protein_id="CAB87133.1"
                             /db_xref="GI:7532758"
                             /translation="MPLTVLAIDDSRTIRELLREALVQAGFEVHLAIDGLDGLEKLEA
                     AKPHAVITDINMPRMDGFGFIRAVREQPQHSALPIIVLTTESAAELKAKAREAGATAW
                     IVKPFDEAKLVSALRRVAVA"
     gene                    10261. .12321
                             /gene="cheA1"
     CDS                     10261. .12321
                             /gene="cheA1"
                             /citation=[1]
                             /codon_start=1
                             /transl_table=11
                             /product="chemotaxis histidine protein kinase"
                             /protein_id="CAB87134.1"
                             /db_xref="GI:7532759"
                             /translation="MDQSDIRSAFFVECEDLMEALNEGLDRIEDTLDDGHDDETVNAV
                     FRAVHSIKGGAGAFKLDALVRFAHQFETTLDALRAGRVSADPPLLALLHKAADRLSDL
                     LQAARTGSETATIDPDDLVAQLAQAAGEEEAGEADAEDLGFVPMRLDLDLPAAAPDEG
```

```
     Query Match              9.1%;   Score 38.8;  DB 1;  Length 14713;
     Best Local Similarity   51.1%;   Pred. No. 16;
     Matches   91;  Conservative   0;  Mismatches   87;  Indels   0;  Gaps   0;

Qy      249 gaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtg 308
                |||   | |||| |    |   |  | |||||  ||    |||| |   || ||||||
Db     3612 GACGAGATCGGCGATCTGATCGCCTCGTTGAACGGCATGGTGACGAAGCTCCGCACGGTG 3671

Qy      309 ctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagctc 368
               || |  |||| |  | |   |   |   ||| |  |||  | | || | || |      |
Db     3672 GTGAACGACGTCGCAAGATCCACGCGCACCGTGGCCGCAGGCGCCGACGAGATGTCCTCC 3731

Qy      369 acgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
               |||  | |||       ||  ||     | ||     ||| || || |||| |||  ||||
Db     3732 ACGGCCGTGAAACTGAGTCAGGGGGCGGCCGAACAGGCCAGCGCCACGCTGCAGGCCT 3789
```

```
RESULT   15
AY047566
LOCUS        AY047566      3314 bp     mRNA                  INV       16-AUG-2001
```

```
DEFINITION   Drosophila melanogaster GH07804 full length cDNA.
ACCESSION    AY047566
VERSION      AY047566.1  GI:15010499
KEYWORDS     FLI_CDNA.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 3314)
  AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
             Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
             Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
             Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
             Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (19-JUL-2001) Berkeley Drosophila Genome Project,
             Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
COMMENT      Sequence submitted by:
             Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory
             Berkeley, CA 94720
             This clone was sequenced as part of a high-throughput process to
             sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
             Science 2000). The sequence has been subjected to integrity checks
             for sequence accuracy, presence of a polyA tail and contiguity
             within 100 kb in the genome. Thus we believe the sequence to
             reflect accurately this particular cDNA clone.  However, there are
             artifacts associated with the generation of cDNA clones that may
             have not been detected in our initial analyses such as internal
             priming, priming from contaminating genomic DNA, retained introns
             due to reverse transcription of unspliced precursor RNAs, and
             reverse transcriptase errors that result in single base changes.
             For further information about this sequence, including its location
             and relationship to other sequences, please visit our Web site
             (http://fruitfly.berkeley.edu) or send email to
             cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..3314
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="99D4-99D5"
                     /clone="GH07804"
     gene            1..3314
                     /gene="CG7921"
                     /note="alignment with genomic scaffold AE003772"
                     /db_xref="FLYBASE:FBgn0039738"
     CDS             612..2438
                     /gene="CG7921"
                     /note="Longest ORF"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0039738"
                     /product="GH07804p"
                     /protein_id="AAK77298.1"
                     /db_xref="GI:15010500"
                     /translation="MSKMRGRILIPLPNTGAMGRKRNNFYMRSLFLLALGIFGLLQYN
```

NFNYLDSRDNVLGDAVTNDSDDAILAMVPATLHKYLTPHSRNHSASGAGALNGAALLL
NASSPGAATASTISFDVYHPPNITEIKRQIVRYNDMQMVLNEDVFGPLQNDSVIIVVQ
VHTRITYLRHLIVSLAQARDISKVLLVFSHDYYDDDINDLVQQIDFCKVMQIFYPYSI
QTHPNEYPGVDPNDCPRNIKKEQALITNCNNAMYPDLYGHYREAKFTQTKHHWIWKAN
RVFNELEVTRYHTGLVLFLEEDHYVAEDFLYLLAMMQQRTKDLCPQCNVLSLGTYLKT
FNYYTYHSKTNKKSYASSLISTNSLLGYNRNNNRNSVQLAVSSSYSSSSSAQSPPPSK
VYVGDTSDNQRASDSSSRRGNLINTVTDTATATAATSDTSNNEQSSHSSSKNGAQTWN
YHVLPSLYSVYQKVEVMPWVSSKHNMGFAFNRTTWSNIRKCARHFCTYDDYNWDWSLQ
HVSQQCLRRKLHAMIVKGPRVFHIGECGVHHKNKNCESNQVISKVQHVLRIARNSHQL
FPRSLTLTVPSLMKKSKLRKGNGGWGDMRDHELCLNMTLATR"
BASE COUNT   928 a   905 c   819 g   662 t
ORIGIN

   Query Match   9.1%;   Score 38.6;   DB 3;   Length 3314;
   Best Local Similarity   51.4%;   Pred. No. 20;
   Matches   89; Conservative   0; Mismatches 84; Indels   0; Gaps   0;

Qy   248 ggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggt 307
            |||| |||| | | | || | | | ||| | || || |||
Db   1043 GGACGTCTTTGGACCGCTGCAGAACGACTCTGTGATAATCGTGGTCCAGGTGCACACGAG 1102

Qy   308 gctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagct 367
            | | || | || ||||| | || ||||| ||| | || | |||||| |
Db   1103 GATCACCTACCTGCGCCACCTGATCGTCAGCCTGGCGCAGGCCCGGGACATTTCGAAGGT 1162

Qy   368 cacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
            | || ||| | || || | | | || | | || |||| ||||||
Db   1163 GCTGCTGGTGTTCTCGCACGACTACTACGACGACGACATCAACGACCTGGTGC 1215

Search completed: February 7, 2002, 11:13:08
Job time: 10314 sec

                       GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:00:41 ; Search time 428.31 Seconds
                                                (without alignments)
                                                852.701 Million cell updates/sec

Title: US-09-394-745-7565
Perfect score: 426
Sequence: 1 gggccgacccacgcgtccag.........:...catcgacacggtgcgagcct 426

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

```
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_1101:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
                 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
                 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
                 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
                 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
                 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
                 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
                 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
                 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
                 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
                 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
                 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
                 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
                 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	179.6	42.2	1356	21	AAC42216	Arabidopsis thalia
2	44	10.3	1766	19	AAV40277	Rat equilibrative
3	37.6	8.8	1929	19	AAV57472	Sorghum bicolor (L
4	36	8.5	2679	13	AAQ28895	Fucose dehydrogena
c 5	35.6	8.4	1433	20	AAX38293	M. tuberculosis se
c 6	33	7.7	1340	21	AAZ45317	DNA encoding a GDP
7	32.8	7.7	4260	9	AAN81768	Sequence encoding
8	32.8	7.7	4380	9	AAN80222	Sequence of Mycoba
9	32.8	7.7	4380	19	AAV05708	Mycobacterium tube
c 10	32.8	7.7	109519	22	AAS08693	Micromonospora DNA
11	32.4	7.6	1032	22	AAD10215	Chimeric moCRE rec
12	32.4	7.6	2346	22	AAD10217	Chimeric recombina
13	32.2	7.6	1470	22	AAF61040	P. putida KT2440-a
14	32.2	7.6	9810	20	AAZ32025	Human METH1 relate
15	32.2	7.6	9810	22	AAC90082	AF018073 cDNA clon
c 16	32	7.5	14327	22	AAF54724	Nucleotide sequenc
17	31.8	7.5	3438	19	AAV39080	S. viridochromogen
18	31.6	7.4	2475	11	AAQ04307	Coding region of h

	19	31.6	7.4	2475	18	AAT61551	Human interleukin-
	20	31.6	7.4	2475	19	AAV38331	Human interleukin-
	21	31.6	7.4	2475	19	AAV17656	Human interleukin-
	22	31.6	7.4	2475	20	AAX22408	Human IL-4 recepto
	23	31.6	7.4	2475	20	AAV08857	Human interleukin-
	24	31.6	7.4	2476	22	AAF57719	Human IL4Ralpha co
	25	31.6	7.4	2478	22	AAF30819	Human interleukin-
	26	31.6	7.4	2520	21	AAA46789	DNA encoding a mam
	27	31.6	7.4	2520	21	AAZ50939	Soluble Interleuki
	28	31.6	7.4	3597	21	AAF20977	Human low adenosin
	29	31.6	7.4	3597	21	AAA34855	Human adenosine re
	30	31.6	7.4	3612	22	AAH98579	Human EST-derived
	31	31.6	7.4	3612	22	AAI57908	Human polynucleoti
	32	31.6	7.4	3612	22	AAI59694	Human polynucleoti
c	33	31.6	7.4	20387	19	AAV62159	HSV-2 strain SB5 C
c	34	31.6	7.4	26338	19	AAV62134	HSV-2 strain SB5 C
c	35	31.6	7.4	117213	19	AAV62176	HSV-2 strain SB5 C
c	36	31.4	7.4	1352	20	AAX89693	Nucleic acid seque
	37	31.4	7.4	5070	21	AAC76230	Human ORFX ORF1785
	38	31.4	7.4	50341	19	AAV22674	DNA sequence of a
	39	31.4	7.4	50341	21	AAZ39519	L5 shuttle phasmid
	40	31.4	7.4	52297	16	AAT51411	Mycobacteriophage
	41	31.4	7.4	52298	14	AAQ47357	L5 mycobacteriopha
	42	31.2	7.3	1125	21	AAA57483	cDNA encoding a ra
	43	31.2	7.3	2679	19	AAV30468	Canine beta-2 adre
	44	31	7.3	375	22	AAF66567	Novel human polynu
c	45	31	7.3	1086	19	AAV47589	Leishmania antigen

ALIGNMENTS

RESULT   1
AAC42216
ID   AAC42216 standard; DNA; 1356 BP.
XX
AC   AAC42216;
XX
DT   17-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 34716.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.

```
PR    23-MAR-1999;    99US-0125788.
PR    25-MAR-1999;    99US-0126264.
PR    29-MAR-1999;    99US-0126785.
PR    01-APR-1999;    99US-0127462.
PR    06-APR-1999;    99US-0128234.
PR    08-APR-1999;    99US-0128714.
PR    16-APR-1999;    99US-0129845.
PR    19-APR-1999;    99US-0130077.
PR    21-APR-1999;    99US-0130449.
PR    23-APR-1999;    99US-0130510.
PR    23-APR-1999;    99US-0130891.
PR    28-APR-1999;    99US-0131449.
PR    30-APR-1999;    99US-0132048.
PR    30-APR-1999;    99US-0132407.
PR    04-MAY-1999;    99US-0132484.
PR    05-MAY-1999;    99US-0132485.
PR    06-MAY-1999;    99US-0132486.
PR    06-MAY-1999;    99US-0132487.
PR    07-MAY-1999;    99US-0132863.
PR    11-MAY-1999;    99US-0134256.
PR    14-MAY-1999;    99US-0134218.
PR    14-MAY-1999;    99US-0134219.
PR    14-MAY-1999;    99US-0134221.
PR    14-MAY-1999;    99US-0134370.
PR    18-MAY-1999;    99US-0134768.
PR    19-MAY-1999;    99US-0134941.
PR    20-MAY-1999;    99US-0135124.
PR    21-MAY-1999;    99US-0135353.
PR    24-MAY-1999;    99US-0135629.
PR    25-MAY-1999;    99US-0136021.
PR    27-MAY-1999;    99US-0136392.
PR    28-MAY-1999;    99US-0136782.
PR    01-JUN-1999;    99US-0137222.
PR    03-JUN-1999;    99US-0137528.
PR    04-JUN-1999;    99US-0137502.
PR    07-JUN-1999;    99US-0137724.
PR    08-JUN-1999;    99US-0138094.
PR    10-JUN-1999;    99US-0138540.
PR    10-JUN-1999;    99US-0138847.
PR    14-JUN-1999;    99US-0139119.
PR    16-JUN-1999;    99US-0139452.
PR    16-JUN-1999;    99US-0139453.
PR    17-JUN-1999;    99US-0139492.
PR    18-JUN-1999;    99US-0139454.
PR    18-JUN-1999;    99US-0139455.
PR    18-JUN-1999;    99US-0139456.
PR    18-JUN-1999;    99US-0139457.
PR    18-JUN-1999;    99US-0139458.
PR    18-JUN-1999;    99US-0139459.
PR    18-JUN-1999;    99US-0139460.
PR    18-JUN-1999;    99US-0139461.
PR    18-JUN-1999;    99US-0139462.
PR    18-JUN-1999;    99US-0139463.
PR    18-JUN-1999;    99US-0139750.
PR    18-JUN-1999;    99US-0139763.
PR    21-JUN-1999;    99US-0139817.
PR    22-JUN-1999;    99US-0139899.
```

```
PR    23-JUN-1999;    99US-0140353.
PR    23-JUN-1999;    99US-0140354.
PR    24-JUN-1999;    99US-0140695.
PR    28-JUN-1999;    99US-0140823.
PR    29-JUN-1999;    99US-0140991.
PR    30-JUN-1999;    99US-0141287.
PR    01-JUL-1999;    99US-0141842.
PR    01-JUL-1999;    99US-0142154.
PR    02-JUL-1999;    99US-0142055.
PR    06-JUL-1999;    99US-0142390.
PR    08-JUL-1999;    99US-0142803.
PR    09-JUL-1999;    99US-0142920.
PR    12-JUL-1999;    99US-0142977.
PR    13-JUL-1999;    99US-0143542.
PR    14-JUL-1999;    99US-0143624.
PR    15-JUL-1999;    99US-0144005.
PR    16-JUL-1999;    99US-0144085.
PR    16-JUL-1999;    99US-0144086.
PR    19-JUL-1999;    99US-0144325.
PR    19-JUL-1999;    99US-0144331.
PR    19-JUL-1999;    99US-0144332.
PR    19-JUL-1999;    99US-0144333.
PR    19-JUL-1999;    99US-0144334.
PR    19-JUL-1999;    99US-0144335.
PR    20-JUL-1999;    99US-0144352.
PR    20-JUL-1999;    99US-0144632.
PR    20-JUL-1999;    99US-0144884.
PR    21-JUL-1999;    99US-0144814.
PR    21-JUL-1999;    99US-0145086.
PR    21-JUL-1999;    99US-0145088.
PR    22-JUL-1999;    99US-0145085.
PR    22-JUL-1999;    99US-0145087.
PR    22-JUL-1999;    99US-0145089.
PR    22-JUL-1999;    99US-0145192.
PR    23-JUL-1999;    99US-0145145.
PR    23-JUL-1999;    99US-0145218.
PR    23-JUL-1999;    99US-0145224.
PR    26-JUL-1999;    99US-0145276.
PR    27-JUL-1999;    99US-0145913.
PR    27-JUL-1999;    99US-0145918.
PR    27-JUL-1999;    99US-0145919.
PR    28-JUL-1999;    99US-0145951.
PR    02-AUG-1999;    99US-0146386.
PR    02-AUG-1999;    99US-0146388.
PR    02-AUG-1999;    99US-0146389.
PR    03-AUG-1999;    99US-0147038.
PR    04-AUG-1999;    99US-0147204.
PR    04-AUG-1999;    99US-0147302.
PR    05-AUG-1999;    99US-0147192.
PR    05-AUG-1999;    99US-0147260.
PR    06-AUG-1999;    99US-0147303.
PR    06-AUG-1999;    99US-0147416.
PR    09-AUG-1999;    99US-0147493.
PR    09-AUG-1999;    99US-0147935.
PR    10-AUG-1999;    99US-0148171.
PR    11-AUG-1999;    99US-0148319.
PR    12-AUG-1999;    99US-0148341.
```

```
PR    13-AUG-1999;    99US-0148565.
PR    13-AUG-1999;    99US-0148684.
PR    16-AUG-1999;    99US-0149368.
PR    17-AUG-1999;    99US-0149175.
PR    18-AUG-1999;    99US-0149426.
PR    20-AUG-1999;    99US-0149722.
PR    20-AUG-1999;    99US-0149723.
PR    20-AUG-1999;    99US-0149929.
PR    23-AUG-1999;    99US-0149902.
PR    23-AUG-1999;    99US-0149930.
PR    25-AUG-1999;    99US-0150566.
PR    26-AUG-1999;    99US-0150884.
PR    27-AUG-1999;    99US-0151065.
PR    27-AUG-1999;    99US-0151066.
PR    27-AUG-1999;    99US-0151080.
PR    30-AUG-1999;    99US-0151303.
PR    31-AUG-1999;    99US-0151438.
PR    01-SEP-1999;    99US-0151930.
PR    07-SEP-1999;    99US-0152363.
PR    10-SEP-1999;    99US-0153070.
PR    13-SEP-1999;    99US-0153758.
PR    15-SEP-1999;    99US-0154018.
PR    16-SEP-1999;    99US-0154039.
PR    20-SEP-1999;    99US-0154779.
PR    22-SEP-1999;    99US-0155139.
PR    23-SEP-1999;    99US-0155486.
PR    24-SEP-1999;    99US-0155659.
PR    28-SEP-1999;    99US-0156458.
PR    29-SEP-1999;    99US-0156596.
PR    04-OCT-1999;    99US-0157117.
PR    05-OCT-1999;    99US-0157753.
PR    06-OCT-1999;    99US-0157865.
PR    07-OCT-1999;    99US-0158029.
PR    08-OCT-1999;    99US-0158232.
PR    12-OCT-1999;    99US-0158369.
PR    13-OCT-1999;    99US-0159293.
PR    13-OCT-1999;    99US-0159294.
PR    13-OCT-1999;    99US-0159295.
PR    14-OCT-1999;    99US-0159329.
PR    14-OCT-1999;    99US-0159330.
PR    14-OCT-1999;    99US-0159331.
PR    14-OCT-1999;    99US-0159637.
PR    14-OCT-1999;    99US-0159638.
PR    18-OCT-1999;    99US-0159584.
PR    21-OCT-1999;    99US-0160741.
PR    21-OCT-1999;    99US-0160767.
PR    21-OCT-1999;    99US-0160768.
PR    21-OCT-1999;    99US-0160770.
PR    21-OCT-1999;    99US-0160814.
PR    21-OCT-1999;    99US-0160815.
PR    22-OCT-1999;    99US-0160980.
PR    22-OCT-1999;    99US-0160981.
PR    22-OCT-1999;    99US-0160989.
PR    25-OCT-1999;    99US-0161404.
PR    25-OCT-1999;    99US-0161405.
PR    25-OCT-1999;    99US-0161406.
PR    26-OCT-1999;    99US-0161359.
```

   Query Match   42.2%;   Score 179.6; DB 21;   Length 1356;
   Best Local Similarity   67.2%;   Pred. No. 9e-44;
   Matches   254; Conservative   0; Mismatches 124; Indels   0; Gaps   0;

Qy   49 atttcaagttcaagcaagagctctggatggtcattagcatgtcctctgttgcggtcgtga 108
           || | | || | || ||||||| | | ||| | || | | | ||||| |
Db   695 atatgagtagcaccgaggaaaaatggatgattggaataatggcttcagctacagtcgtca 754

Qy   109 agttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgccc 168
           |||| | ||||||||| |||||| | | ||| |||| || || || ||||||||| || |
Db   755 agtttctgctcatgctttactgcaggagtttccagaacgaaattgtcagggcctatgcac 814

Qy   169 aggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctg 228
           | || || |||| || || ||||| || || || ||||| | | | || | | ||||
Db   815 aagatcacctctttgatgttatcaccaattcagtcggtttagcaaccgctgttttagctg 874

Qy   229 tccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgatca 288
           | | | | ||||||||| || || ||| || ||||| || || |||| || ||||
Db   875 taaaattctactggtggattgatccctctggggctatactaattgccctgtatacaatca 934

Qy   289 cgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccgg 348
            || |||||| |||| || || ||||| || |||||||||| || || ||
Db   935 gcacatgggcaagaacagttctagagaatgtccattcactgataggacgctcagcaccac 994

Qy   349 cagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcaca 408
           |||| | | | ||| || || |||| |||||| |||||||| || ||| |||| | | ||||
Db   995 cagatttcttggcgaaactaacgttcttgatttggaaccatcacgagaagataaaacaca 1054

Qy   409 tcgacacggtgcgagcct 426
           | |||||| ||| ||||||
Db   1055 tagacacagtgagagcct 1072

RESULT   2
AAV40277
ID   AAV40277 standard; cDNA; 1766 BP.
XX
AC   AAV40277;
XX
DT   13-OCT-1998   (first entry)
XX
DE   Rat equilibrative nucleoside transporter 1 encoding cDNA.
XX
KW   Rat; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;
KW   rENT2; coronary; cerebrovascular anoxia; viral infection; cancer; ss.
XX
OS   Rattus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   5..1378

    Query Match   10.3%;   Score 44; DB 19; Length 1766;
    Best Local Similarity   49.2%; Pred. No. 0.0024;
    Matches 116; Conservative   0; Mismatches 120; Indels   0; Gaps   0;

Qy   34 atcaggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgtcct 93
            I II I II I   I I   II IIIII II IIIII I III   IIII II
Db   1152 agcaccactacctgccctccctctttaagcatgatgtctggttcatcaccttcatggccg 1211

Qy   94 ctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgagatcg 153
            I IIII II II   II IIII   I IIII   III I   III I
Db   1212 cctttgccttctccaatggctacctcgccagcctctgcatgtgcttcgggcccaagaaag 1271

Qy   154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
            I I I   I   II   I I   II I II I   II I I I IIIII I

```
Db    1272  tcaaaccggctgaggcagagactgccggaaacatcatgtccttctttctgtgtctgggcc 1331

Qy     214  cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactg 269
                 ||| |||    ||||| || |  | || | || | ||| | || | |||
Db    1332  tggctctgggagctgtgttgtccttcttgttaagggcacttgtgtgagcgaccctg 1387
```

RESULT   3
AAV57472
```
ID    AAV57472 standard; cDNA; 1929 BP.
XX
AC    AAV57472;
XX
DT    14-DEC-1998   (first entry)
XX
DE    Sorghum bicolor (L.) Moench cytochrome P450ox monooxygenase cDNA.
XX
KW    Cytochrome P450 monooxygenase; P450ox; Sorghum bicolor (L.) Moench;
KW    Sinapis alba; biosynthetic conversion; aldoxime; nitrile; cyanohydrin;
KW    cyanogenic glycoside; transgenic plant; resistance; ds.
XX
OS    Sorghum bicolor.
XX
FH    Key               Location/Qualifiers
FT    CDS               81..1676
FT                      /*tag=  a
FT                      /product= "cytochrome P450 monooxygenase"
XX
PN    WO9840470-A2.
XX
PD    17-SEP-1998.
XX
PF    05-MAR-1998;    98WO-EP01253.
XX
PR    08-DEC-1997;    97EP-0810954.
PR    07-MAR-1997;    97EP-0810132.
XX
PA    (NOVS ) NOVARTIS AG.
PA    (UYRO-) UNIV ROYAL VETERINARY & AGRIC.
XX
PI    Bak S,  Halkier BA,  Kahn RA,  Moeller BL;
XX
DR    WPI; 1998-520808/44.
DR    P-PSDB; AAW79067.
XX
PT    Cytochrome P450 monooxygenase of the cyanogenic glycoside pathway -
PT    useful for the production of plants with improved nutritive value or
PT    pest resistance
XX
PS    Example 6; Page 41-43; 32pp; English.
XX
CC    The present sequence encodes a cytochrome P450 monooxygenase from
CC    Sorghum bicolor (L.) Moench, designated P450ox. Cytochrome P450
CC    monooxygenase catalyses: (i) the conversion of aldoxime to a nitrile;
CC    and (ii) the nitrile to the corresponding cyanohydrin. DNA encoding
CC    cytochrome P450 monooxygenase can be used to obtain transgenic plants,
CC    for the purpose of improving the nutritive value or pest resistance of
```

```
CC    the plant. Cytochrome P450 monooxygenase catalyses the conversion of
CC    aldoximes to nitriles to cyanohydrins, which are the precursors of toxic
CC    cyanogenic glycosides, so staple food such as cassava and lima beans,
CC    as well as animal feed such as white clover, can be rendered less toxic
CC    by blocking the cytochrome P450 monooxygenase activity. Introducing the
CC    enzyme to plants or to certain tissues could help reduce crop damage
CC    since the product is also toxic to insects, acarids and nematodes.
XX
SQ    Sequence 1929 BP; 374 A; 683 C; 577 G; 295 T; 0 other;


     Query Match              8.8%;   Score 37.6;  DB 19;   Length 1929;
     Best Local Similarity   49.5%;   Pred. No. 0.2;
     Matches   97; Conservative    0; Mismatches   99; Indels    0; Gaps    0;


Qy     140 caagaatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactc 199
           || | |   ||| | ||  || |||| |||||  ||| | |||    |      |
Db     779 catggacatgatggccagcttctccgccgaggacttcttccccaacgccgccggccgcct 838

Qy     200 tgtcggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttgg 259
            | || ||   |||||||  | |||||  ||| |           |  |  || | | |
Db     839 cgccgaccgcctctcgggcttcctcgcccgccgcgagcgcatcttcaacgagctcgacgt 898

Qy     260 cgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
            |  | |   || |   |  | || | || |      ||||||  | |||| ||| ||||
Db     899 cttcttcgagaaggtcatcgaccagcacatggacccggcgcgcccgtgccggacaacgg 958

Qy     320 aggcacactgataggc 335
            |||   || | | |
Db     959 cggcgacctcgtcgac 974


RESULT   4
AAQ28895
ID    AAQ28895 standard; DNA; 2679 BP.
XX
AC    AAQ28895;
XX
DT    01-MAR-1993  (first entry)
XX
DE    Fucose dehydrogenase DNA.
XX
KW    Arthrobacter oxidans; F1; induction; assay; ss.
XX
OS    Arthrobacter oxidans F1.
XX
FH    Key             Location/Qualifiers
FT    CDS             844..1809
FT                    /*tag=  a
XX
PN    EP506262-A.
XX
PD    30-SEP-1992.
XX
PF    13-MAR-1992;   92EP-0302170.
XX
PR    29-MAR-1991;   91JP-0089184.
```

PA   (TAKI) TAKARA SHUZO CO LTD.
XX
PI   Kato I, Kotani H, Mitta M, Sakai T;
XX
DR   WPI; 1992-325548/40.
DR   P-PSDB; AAR27118.
XX
PT   Isolated gene encoding L-fucose dehydrogenase - useful for prodn.
PT   of enzyme by genetic engineering
XX
PS   Claim 1; Page 8; 16pp; English.
XX
CC   Genomic DNA from Arthrobacter oxidans F1 was subjected to
CC   restriction enzyme analysis and the N-terminal amino acid sequence
CC   of L-fucose dehydrogenase determined. A degenerate probe was
CC   synthesised based on this amino acid sequence. The probe was used
CC   to screen an Arthrobacter cDNA library to isolate a L-fucose dehydro-
CC   genase clone. The isolation of such a clone provides a convenient
CC   method for prodn. of L-fucose dehydrogenase without the need for
CC   induction by L-fucose. The probe may be used to evaluate the extent
CC   of expression of L-fucose dehydrogenase. The DNA sequence is
CC   widely used to assay L-fucose levels.
CC   See also AAQ28894.
XX
SQ   Sequence 2679 BP; 481 A; 917 C; 868 G; 413 T; 0 other;

    Query Match   8.5%;   Score 36;   DB 13;   Length 2679;
    Best Local Similarity   47.0%;   Pred. No. 0.67;
    Matches 111; Conservative   0; Mismatches 125; Indels   0; Gaps   0;

Qy   37 aggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgtcctctg 96
            ||||` || || ||| | || | | || ||| |||| | ||| |
Db   1130 aggacaccgagggcttcgacgtcccggacgacctcatccgggtccgcgactactcccgcg 1189

Qy   97 ttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtga 156
            | ||| || | | || | || | | | | |||||| |
Db   1190 acggggtgctgcgctccatcgaggaaagcctgcagcggctggggaccgaccggatcgaca 1249

Qy   157 gggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcgg 216
            | ||||| ||| |||| | | | || | | | | ||| | | |||
Db   1250 tcgtctacatccacgaccctgacgactactggaccgaggccgtggagggcgccgccccgg 1309

Qy   217 cgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatc 272
            ||||| ||| | ||| || || || || | ||||| ||| |
Db   1310 cgctgtccgccctgcgggacgaaggggtcatcagggcctggggcgcaggcatgaac 1365

RESULT   5
AAX38293/c
ID   AAX38293 standard; DNA; 1433 BP.
XX
AC   AAX38293;
XX
DT   16-JUN-1999 (first entry)
XX

  Query Match   8.4%;   Score 35.6;   DB 20;   Length 1433;
  Best Local Similarity   46.7%;   Pred. No. 0.7;
  Matches 113; Conservative   0; Mismatches 129; Indels   0; Gaps   0;

Qy   91 cctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgaga 150
           || | |||| | | | || || |||| | | || || || | |
Db   1102 CCGACGGTGCGTTTGCGTAGATCGACCTCGTAGTGGACGTCCAAAACCAGCCGCGGCACG 1043

Qy   151 tcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctgg 210
            | |||| || | | |||| || | | | | | | || || || ||
Db   1042 CCAACCGGGCGAACTCGGTGAACCAGTTGGAGGCGGGCGCCCCGGGCGGAGATGGGGGGG 983

Qy   211 tctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactga 270
             | ||| | || || | ||| | | | |||||| || ||
Db   982 GGGCCCCGCCCTCGTCGATCAGGATGGAATCGACCTTCGTCGACAATGGCGTAATGGTGC 923

Qy   271 tcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtaggcacactga 330
            ||| || || ||||| ||| | ||| || | | | | |||| | | || |

```
Db      922 CCGCGCTGCACCAGATCATCCAGTGAGTGCGCCATGTTGTCGCGCAGGTAGTCGAACCCA 863

Qy      331 ta 332
                |
Db      862 AA 861


RESULT   6
AAZ45317/c
ID      AAZ45317 standard; DNA; 1340 BP.
XX
AC      AAZ45317;
XX
DT      27-MAR-2000  (first entry)
XX
DE      DNA encoding a GDP-4-keto-6-deoxy-D-mannose epimerase/reductase.
XX
KW      GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;
KW      GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;
KW      ascorbic acid pathway enzyme; hexokinase; glucose phosphate isomerase;
KW      phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;
KW      GDP-D-mannose pyrophosphorylase; GDP-D-mannose:GDP-L-galactose epimerase;
KW      GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;
KW      L-galactono-gamma-lactone dehydrogenase; ester; ss.
XX
OS      Homo sapiens.
XX
FH      Key             Location/Qualifiers
FT      CDS             75..1040
FT                      /*tag=  a
FT                      /product= "GDP-4-keto-6-deoxy-D-mannose epimerase/
FT                                reductase"
XX
PN      WO9964618-A1.
XX
PD      16-DEC-1999.
XX
PF      26-MAY-1999;    99WO-US11576.
XX
PR      08-JUN-1998;    98US-0088549.
PR      17-MAR-1999;    99US-0125073.
PR      18-MAR-1999;    99US-0125054.
XX
PA      (DCVB-) DCV INC DBA BIO-TECH RESOURCES.
XX
PI      Berry A,  Running JA,  Severson DK,  Burlingame RP;
XX
DR      WPI; 2000-105890/09.
DR      P-PSDB; AAY54116.
XX
PT      Production of ascorbic acid or esters, using microorganisms or plants
PT      which have genetic modification in enzymes involved in the ascorbic
PT      acid synthesis pathway  -
XX
PS      Claim 26; Page 171-173; 187pp; English.
XX
CC      The present sequence encodes a GDP-4-keto-6-deoxy-D-mannose epimerase/
```

```
CC    reductase. The enzyme catalyses the conversion of GDP-D-mannose to
CC    GDP-L-galactose. The enzyme can be modified, and used to produce
CC    transgenic microorganisms, which can be used in fermentation techniques
CC    to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is
CC    modified to increase its action. Other ascorbic acid pathway enzymes
CC    which may be used in the method of the invention include hexokinases,
CC    glucose phosphate isomerases, phosphomannose isomerases,
CC    phosphomannomutases, GDP-D-mannose pyrophosphorylases,
CC    GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,
CC    L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and
CC    L-galactono-gamma-lactone dehydrogenases. The methods can be used for
CC    the production of ascorbic acid or esters using microorganisms or plants.
XX
SQ    Sequence 1340 BP; 311 A; 400 C; 376 G; 253 T; 0 other;


   Query Match               7.7%;   Score 33;   DB 21;   Length 1340;
   Best Local Similarity   55.8%;   Pred. No. 4;
   Matches   63;   Conservative   0;   Mismatches   50;   Indels   0;   Gaps   0;


Qy     203 cggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgc 262
               || |   || |   |  | |||||||||    |  | |||  ||   |  ||||||||   || |    || |
Db     692 CGTCAGGGCCGAGCCGCTGCTCTTGGCCAGGTGCACCTTGTGGATGAGGCCAGGCAGCAC 633


Qy     263 catactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggaga 315
                 |     | || |||  | | |   || | |   || ||||  ||  ||| ||
Db     632 GTGGCCATCCTCGATGTTGAAGTTGTCGTGGGGCCCGAAAACGTTGGTGGGGA 580



RESULT    7
AAN81768
ID    AAN81768 standard; DNA; 4260 BP.
XX
AC    AAN81768;
XX
DT    29-DEC-1990   (first entry)
XX
DE    Sequence encoding Mycobacterium tuberculosis 540 and 517 AA residue
DE    proteins.
XX
KW    Diagnosis; assay; M.bovis; vaccine; ds.
XX
OS    Mycobacterium tuberculosis.
XX
FH    Key             Location/Qualifiers
FT    CDS             252..1874
FT                    /*tag=  a
FT                    /label=540 AA protein
FT                    /note="AAP81351"
FT    CDS              complement (3948..2395)
FT                    /*tag=  b
FT                    /label=517 AA protein
FT                    /note="AAP81868"
XX
PN    WO8806591-A.
XX
PD    07-SEP-1988.
```

XX
PF   25-FEB-1988;   88WO-US00598.
XX
PR   24-FEB-1988;   88US-0159667.
PR   06-FEB-1987;   87US-0019529.
XX
PA   (SCRI-) SCRIPPS CLINIC & RE.
XX
PI   Shinnick T, Houghten R;
XX
DR   WPI; 1988-271136/38.
DR   P-PSDB; AAP81351, AAP81868.
XX
PT   Recombinant mycobacterial peptide(s) -
PT   used in assays for diagnosis of infection, for producing
PT   vaccines and for producing antibodies
XX
PS   Disclosure; Fig 2a-2d; 116pp; English.
XX
CC   An isolated DNA molecule that consists essentially of the nucleotide
CC   sequence that corresponds to the sequence represented by position 3950
CC   to about 2390 and from position 3948 through position 2398 of AAN81768
CC   is claimed. Also claimed is a peptide sequence that consists of a 5-40
CC   AA residue sequence that corresponds to a sequence of the 540 AA residue
CC   protein (AAP81351) or the 517 AA residue protein (AAP81868) coded for by
CC   the DNA sequence. The proteins can be used for determining previous
CC   immunological exposure of a mammal to M.tuberculosis or M.bovis and
CC   for producing a vaccine.
XX
SQ   Sequence 4260 BP; 733 A; 1332 C; 1481 G; 714 T; 0 other;

    Query Match   7.7%; Score 32.8; DB 9; Length 4260;
    Best Local Similarity   56.5%; Pred. No. 7.2;
    Matches   61; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

Qy   154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
               ||||||| || | || | || | ||| | || ||| | | | || | || |
Db   3901 tgagggtctgccacctgccccgtaatgtcgctggtatggcaagcaccgacgccgcggccc 3960

Qy   214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcg 261
               | | ||||| | | | || || ||| ||| || | ||| | ||
Db   3961 aagagttgctccgcgacgcgttcacccggttgatcgaacatgtcgacg 4008

RESULT   8
AAN80222
ID   AAN80222 standard; DNA; 4380 BP.
XX
AC   AAN80222;
XX
DT   19-MAR-1991 (first entry)
XX
DE   Sequence of Mycobacterium tuberculosis DNA contg. gene encoding 65
DE   protein.
XX
KW   Antigen; vaccine; ds.

```
XX                        .
OS    Mycobacterium tuberculosis.
XX
FH    Key               Location/Qualifiers
FT    CDS               192..1874
FT                      /*tag=  a
FT    CDS               complement (2398..4101)
FT                      /*tag=  b
XX
PN    WO8805823-A.
XX
PD    11-AUG-1988.
XX
PF    01-FEB-1988;    88WO-US00281.
XX
PR    02-FEB-1987;    87US-0010007.
XX
PA    (WHIT-) WHITEHEAD INST BIOM.
XX
PI    Husson RN,  Young RA,  Shinnick TM;
XX
DR    WPI; 1988-235175/33.
DR    P-PSDB; AAP80215, AAP80216.
XX
PT    Genes encoding Mycobacterium tuberculosis protein antigens -
PT    useful for developing reagents for diagnosis, prevention and
PT    treatment of tuberculosis
XX
PS    Claim 12; Fig 8; 82pp; English.
XX
CC    The gene was isolated by probing a lambda gt11 expression library of
CC    M.tuberculosis DNA with monoclonal antibodies directed against
CC    M.tuberculosis-specific antigens. The 19kD, 71kD and the 65kD proteins
CC    and genes are claimed, and so is a vaccine comprising DNA encoding
CC    M.tuberculosis protein in a recombinant vaccine vector. AAP80216 is
CC    encoded on the complementary strand.
XX
SQ    Sequence 4380 BP; 757 A; 1373 C; 1512 G; 738 T; 0 other;


   Query Match                7.7%;   Score 32.8;  DB 9;  Length 4380;
   Best Local Similarity   56.5%;   Pred. No. 7.3;
   Matches   61;  Conservative   0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
            ||||||| || |  || |  || |    ||| |  || ||| | | || | | || |
Db   4021 tgagggtctgccacctgccccgtaatgtcgctggtatggcaagcaccgacgccgcggccc 4080


Qy    214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcg 261
            | | ||||| |  | | || || ||| ||| || | ||| | ||
Db   4081 aagagttgctccgcgacgcgttcacccggttgatcgaacatgtcgacg 4128



RESULT    9
AAV05708
ID    AAV05708 standard; DNA; 4380 BP.
XX
```

CC   This DNA sequence encodes the 65 kDa heat shock protein (see
CC   AAW44702), designated Mt Hsp65, of Mycobacterium tuberculosis. This
CC   protein cross-reacts with a component of articular cartilage, human
CC   Hsp60, that is up-regulated in the joints of arthritic patients. A
CC   claimed method for treating or preventing an autoimmune disease in
CC   a mammal comprises: (a) providing a particle coated with an antigen
CC   against which an immune response is mounted in the autoimmune
CC   disease; (b) delivering the particle into the recipient cell of the
CC   mammal; and (c) repeating step (b) until either a reduction in a
CC   cytotoxic immune response or a desensitizing immune response is
CC   induced in the mammal. Alternatively, step (a) comprises providing
CC   a nucleic acid construct comprising a coding sequence for the
CC   antigen, operably linked to control elements such that the coding
CC   sequence can be transcribed and translated in a recipient cell, and
CC   delivering the construct to the recipient cell using a gene gun.
CC   The antigen of step (a) is selected from collagen, Mt Hsp65,
CC   myelin basic protein, myelin oligodendrocyte glycoprotein,
CC   proteolipid protein, and epitopes thereof. These antigens mitigate
CC   cytotoxic responses and elicit antigen desensitisation. The method

CC   is used especially for treating rheumatoid arthritis or multiple
CC   sclerosis. It represents a novel use for the known Mt Hsp65 gene.
XX
SQ   Sequence 4380 BP; 757 A; 1371 C; 1514 G; 738 T; 0 other;

  Query Match   7.7%;   Score 32.8; DB 19;   Length 4380;
  Best Local Similarity   56.5%;   Pred. No. 7.3;
  Matches   61; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

Qy   154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
           ||||||| || | || | ·|| | ||| | || ||| | | | || | || |
Db   4021 tgagggtctgccacctgccccgtaatgtcgctggtatggcaagcaccgacgccgcggccc 4080

Qy   214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcg 261
           | | |||||| | | || || ||| ||| || | ||| | ||
Db   4081 aagagttgctccgcgacgcgttcacccggttgatcgaacatgtcgacg 4128

RESULT 10
AAS08693/c
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
XX
DT   26-SEP-2001   (first entry)
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW   Everinomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key   Location/Qualifiers
FT   CDS   complement (132..1382)
FT   /*tag= a
FT   /product= "EvdA"
FT   RBS   complement (1389..1394)
FT   /*tag= b
FT   CDS   complement (1490..2611)
FT   /*tag= c
FT   /product= "EvdB"
FT   RBS   complement (2618..2622)
FT   /*tag= d
FT   CDS   complement (2622..3860)
FT   /*tag= e
FT   /product= "EvdC"
FT   RBS   complement (3867..3870)
FT   /*tag= f
FT   CDS   4143..5312
FT   /*tag= g
FT   /product= "EvdD"
FT   RBS   4134..4138
FT   /*tag= h
FT   CDS   5309..6235
FT   /*tag= i

```
FT                          /product= "EvdE"
FT      CDS                 6232..7275
FT                          /*tag=  j
FT                          /product= "EvdF"
FT      RBS                 6226..6229
FT                          /*tag=  k
FT      CDS                 7272..8327
FT                          /*tag=  l
FT                          /product= "EvdG"
FT      CDS                 8342..9364
FT                          /*tag=  m
FT                          /product= "EvdH"
FT      RBS                 8333..8336
FT                          /*tag=  n
FT      CDS                 complement (9463..10224)
FT                          /*tag=  o
FT                          /product= "EvdI"
FT      RBS                 complement (10232..10235)
FT                          /*tag=  p
FT      CDS                 10424..11176
FT                          /*tag=  q
FT                          /product= "EvdJ"
FT      CDS                 12027..12455
FT                          /*tag=  r
FT                          /product= "EvdK"
FT                          /partial
FT                          /note= "No start codon"
FT      CDS                 complement (12108..13022)
FT                          /*tag=  s
FT                          /product= "EvdL"
FT      RBS                 complement (13027..13030)
FT                          /*tag=  t
FT      CDS                 complement (14410..15363)
FT                          /*tag=  u
FT                          /product= "EvrA"
FT      RBS                 complement (15369..15373)
FT                          /*tag=  v
FT      CDS                 complement (15380..16414)
FT                          /*tag=  w
FT                          /product= "EvrB"
FT      CDS                 complement 16419..17873
FT                          /*tag=  x
FT                          /product= "EvrC"
FT      CDS                 complement (17870..18934)
FT                          /*tag=  y
FT                          /product= "EvrD"
FT      CDS                 19374..20906
FT                          /*tag=  z
FT                          /product= "EvrE"
FT      CDS                 21064..22542
FT                          /*tag=  aa
FT                          /product= "EvrF"
FT      RBS                 21056..22542
FT                          /*tag=  ab
FT      CDS                 22748..24172
FT                          /*tag=  ac
FT                          /product= "EvrG"
```

```
FT     RBS             22736..22740
FT                     /*tag=  ad
FT     CDS             complement (24177..25223)
FT                     /*tag=  ae
FT                     /product= "EvrH"
FT     RBS             complement (25230..25233)
FT                     /*tag=  af
FT     CDS             25550..26626
FT                     /*tag=  ag
FT                     /product= "EvrI"
FT     CDS             26685..30479
FT                     /*tag=  ah
FT                     /product= "EvrJ"
FT     RBS             26672..26676
FT                     /*tag=  ai
FT     CDS             complement (30557..31876)
FT                     /*tag=  aj
FT                     /product= "EvrK"
FT     RBS             complement (31885..31888)
FT                     /*tag=  ak
FT     CDS             complement (31941..32882)
FT                     /*tag=  al
FT                     /product= "EvrL"
FT     CDS             complement (33167..34405)
FT                     /*tag=  am
FT                     /product= "EvrM"
FT     RBS             complement (34414..34418
FT                     /*tag=  an
FT     CDS             complement (34449..35210)
FT                     /*tag=  ao
FT                     /product= "EvrN"
FT     RBS             complement (35219..35221)
FT                     /*tag=  ap
FT     CDS             complement (35294..36238)
FT                     /*tag=  aq
FT                     /product= "EvrO"
FT     CDS             complement (36235..36963)
FT                     /*tag=  ar
FT                     /product= "EvrP"
FT     CDS             complement (36998..38026)
FT                     /*tag=  as
FT                     /product= "EvrQ"
FT     CDS             complement (38072..38566)
FT                     /*tag=  at
FT                     /product= "EvrR"
FT     CDS             complement (38892..40163)
FT                     /*tag=  au
FT                     /product= "EvrS"
FT     CDS             complement (40216..40890)
FT                     /*tag=  av
FT                     /product= "EvrT"
FT     RBS             complement (40899..40902)
FT                     /*tag=  aw
FT     CDS             complement (40887..41576)
FT                     /*tag=  ax
FT                     /product= "EvrU"
FT     CDS             complement (41679..42707)
```

```
FT                       /*tag=  ay
FT                       /product= "EvrV"
FT      RBS              complement (42714..42717)
FT                       /*tag=  az
FT      CDS              complement (42810..43799)
FT                       /*tag=  ba
FT                       /product= "EvrW"
FT      RBS              complement (43807..43811)
FT                       /*tag=  bb
FT      CDS              complement (43799..44866)
FT                       /*tag=  bc
FT                       /product= "EvrX"
FT      CDS              complement (45014..45760)
FT                       /*tag=  bd
FT                       /product= "EvrY"
FT      RBS              complement (45767..45770)
FT                       /*tag=  be
FT      CDS              complement (45962..46714)
FT                       /*tag=  bf
FT                       /product= "EvrZ"
FT      RBS              complement (45952..45956)
FT                       /*tag=  bg
FT      CDS              complement (47156..49234)
FT                       /*tag=  bh
FT                       /product= "EvsA"
FT      CDS              51627..52715
FT                       /*tag=  bi
FT                       /product= "EvsB"
FT      RBS              51629..51622
FT                       /*tag=  bj
FT      CDS              52889..53557
FT                       /*tag=  bk
FT                       /product= "EvsC"
FT      CDS              53554..54207
FT                       /*tag=  bl
FT                       /product= "EvbA"
FT      CDS              complement (54362..55117)
FT                       /*tag=  bm
FT                       /product= "EvbB"
FT      RBS              complement (55125..55128)
FT                       /*tag=  bn
FT      CDS              complement (55135..56094)
FT                       /*tag=  bo
FT                       /product= "EvbC"
FT      RBS              complement (56100..56103)
FT                       /*tag=  bp
FT      CDS              complement (56184..56813)
FT                       /*tag=  bq
FT                       /product= "EvbC2"
FT      CDS              56961..58709
```

```
   Query Match             7.7%;   Score 32.8;   DB 22;   Length 109519;
   Best Local Similarity   52.1%;   Pred. No. 25;
   Matches   73;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    153 gtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtc 212
              | || ||| | | | ||| |   ||| || ||   ||| || |    |||| |||||
```

```
Db    3497  GCGATGGCGGAGGGCGAGGCCGCCGTCTGCGGCGCGCTCAAGGACGCCCCCGGCGTGGTC 3438

Qy     213  tcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatc 272
                | | |||||   | | | |   |   | ||| ||    | |   ||    | ||| ||
Db    3437  ACCGAGCTGCATTCCGACGGCGCCGGCGGCTGGCTGCTGTCGGGCCGCAAGGTGCTGGTC 3378

Qy     273  gcgttgtacacgatcacgac 292
                ||     | ||| ||||
Db    3377  AGCATGGCGCCCATCGCGAC 3358


RESULT  11
AAD10215
ID    AAD10215 standard; DNA; 1032 BP.
XX
AC    AAD10215;
XX
DT    24-SEP-2001  (first entry)
XX
DE    Chimeric moCRE recombinase DNA.
XX
KW    Maize; site specific recombinase; expression cassette; chimeric; moCRE;
KW    Cre protein; ds.
XX
OS    Chimeric - Zea mays.
OS    Chimeric - Bacteriophage P1.
XX
FH    Key             Location/Qualifiers
FT    CDS             1..1032
FT                    /*tag=  a
FT                    /product= "Chimeric moCRE protein"
XX
PN    US6262341-B1.
XX
PD    17-JUL-2001.
XX
PF    17-NOV-1998;   98US-0193503.
XX
PR    18-NOV-1997;   97US-0065613.
PR    18-NOV-1997;   97US-0065627.
PR    08-SEP-1998;   98US-0099435.
XX
PA    (PION-) PIONEER HI-BRED INT INC.
XX
PI    Baszczynski CL,  Lyznik LA,  Gordon-Kamm WJ,  Guan X,   Rao AG;
PI    Tagliani LA;
XX
DR    WPI; 2001-450495/48.
DR    P-PSDB; AAE05410.'
XX
PT    Integrating DNA of interest into genome of eukaryotic cell, by
PT    transforming plant cell with transfer cassette comprising DNA flanked
PT    by target sites for site-specific recombinases and providing
PT    recombinases in cell  -
XX
PS    Disclosure; Column 15-16; 30pp; English.
XX
```

CC   The invention relates to compositions and methods for introducing
CC   a DNA of interest into a genomic target site. The methods and
CC   compositions involve the use of a combination of target sites for two
CC   site specific recombinases and expression of a chimeric recombinase
CC   with dual target site specificity. The compositions comprise novel
CC   site-specific recombinases with specificities to multiple target sites,
CC   and nucleotide sequences and expression cassettes encoding these
CC   recombinases or target sites. The method of integrating foreign DNA
CC   into genome of eukaryotic cell involves transforming the cell having
CC   target sites for the novel recombinase with a DNA of interest that is
CC   flanked by corresponding target sites. The method is useful for
CC   constructing stably transformed eukaryotic cells, preferably plant
CC   cells. The present sequence is a chimeric recombinase DNA encoding
CC   moCRE, Cre protein from Bacteriophage P1 with maize preferred codons.
XX
SQ   Sequence 1032 BP; 228 A; 326 C; 301 G; 177 T; 0 other;

    Query Match   7.6%;   Score 32.4; DB 22; Length 1032;
    Best Local Similarity   54.1%;   Pred. No. 5.5;
    Matches   66; Conservative   0; Mismatches   56; Indels   0; Gaps   0;

Qy   299 gcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacct 358
               || |||||| | |||||| | | || | || | || || || || | || |
Db   12 gctcacggttcaccagaaccttccggctcttccagtggacgcgacgtccgatgaagtcag 71

Qy   359 gacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggt 418
               || ||| |||| | || || || || | || || || | | || ||||||| | | | ||
Db   72 gaagaacctcatggacatgttccgcgacaggcaagcgttcagcgagcacacctggaagat 131

Qy   419 gc 420
               ||
Db   132 gc 133

RESULT 12
AAD10217
ID   AAD10217 standard; DNA; 2346 BP.
XX
AC   AAD10217;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Chimeric recombinase DNA encoding moCre:FLPm protein.
XX
KW   Site specific recombinase; expression cassette; chimeric;
KW   moCre:FLPm protein; ds.
XX
OS   Chimeric - Saccharomyces sp.
OS   Chimeric - Bacteriophage P1.
OS   Chimeric - Zea mays.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2346
FT   /*tag= a
FT   /product= "Chimeric moCre:FLPm protein"
XX

CC   The invention relates to compositions and methods for introducing
CC   a DNA of interest into a genomic target site. The methods and
CC   compositions involve the use of a combination of target sites for two
CC   site specific recombinases and expression of a chimeric recombinase
CC   with dual target site specificity. The compositions comprise novel
CC   site-specific recombinases with specificities to multiple target sites,
CC   and nucleotide sequences and expression cassettes encoding these
CC   recombinases or target sites. The method of integrating foreign DNA
CC   into genome of eukaryotic cell involves transforming the cell having
CC   target sites for the novel recombinase with a DNA of interest that is
CC   flanked by corresponding target sites. The method is useful for
CC   constructing stably transformed eukaryotic cells, preferably plant
CC   cells. The present sequence is a chimeric recombinase DNA encoding
CC   moCre:FLPm, Cre protein from Bacteriophage P1 and FLP from
CC   Saccharomyces, both maize preferred codons.
XX
SQ   Sequence 2346 BP; 534 A; 807 C; 599 G; 406 T; 0 other;

      Query Match   7.6%; Score 32.4; DB 22; Length 2346;
      Best Local Similarity   54.1%; Pred. No. 7.6;
      Matches   66; Conservative   0; Mismatches   56; Indels   0; Gaps   0;

Qy   299 gcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacct 358
            || |||||| | |||||| | | || || | || || | | || |
Db   12 gctcacggttcaccagaaccttccggctcttccagtggacgcgacgtccgatgaagtcag 71

Qy   359 gacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggt 418
            || ||| |||| | || || || | || || || | | | | |||||| | | | |
Db   72 gaagaacctcatggacatgttccgcgacaggcaagcgttcagcgagcacacctggaagat 131

Qy   419 gc 420

RESULT 13
AAF61040
ID   AAF61040 standard; DNA; 1470 BP.
XX
AC   AAF61040;
XX
DT   16-MAY-2001 (first entry)
XX
DE   P. putida KT2440-associated DNA ORF06499.
XX
KW   Transgenic plant; detection; probe; amplification; vaccine carrier;
KW   microbial production strain; biological remediation; ds.
XX
OS   Pseudomonas putida.
XX
PN   DE19935088-A1.
XX
PD   01-FEB-2001.
XX
PF   27-JUL-1999;   99DE-1035088.
XX
PR   27-JUL-1999;   99DE-1035088.
XX
PA   (TIGR-) TIGR INST GENOMIC RES.
PA   (QUIA-) QUIAGEN GMBH.
PA   (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA   (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
DR   WPI; 2001-192469/20.
XX
PT   New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT   safe genetic engineering host, allow detection in presence of other
PT   related bacteria -
XX
PS   Claim 1a; Page 90-91; 158pp; German.
XX
CC   This invention describes novel DNA sequences (I) for specific detection
CC   of Pseudomonas putida KT2440. The invention also describes (1)
CC   recombinant expression vector containing (I); (2) prokaryotic or
CC   eukaryotic cells transformed or transfected with (I) or the vector of
CC   (1); (3) production of expression products by culturing cells of (2);
CC   (4) expression products, or their fragments, of (I) and synthetic
CC   proteins or peptides with the same sequences (A); (5) poly- or
CC   mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC   hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC   plants that contain transformed or transfected cells of (2); (8)
CC   detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
CC   carrying one or more (I). (I), and their fragments, are used as probes
CC   to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC   polymerase chain reaction, and for production of transgenic plants. (I),
CC   or antibodies that recognize their expression products, are used for
CC   detecting the presence of KT2440, particularly in presence of other,

CC   even closely related, bacteria. KT2440 is one of the bacteria classified
CC   as safe, by the National Institutes of Health, for genetic engineering
CC   work, e.g. as microbial production strains, for biological remediation
CC   and as vaccine carriers. (I) are exclusive to KT2440 with no significant
CC   homology with sequences in other bacteria (specifically the closely
CC   related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC   has greater catabolic activity and better survival in, and adaptation to,
CC   the rhizosphere and soil.
XX
SQ   Sequence 1470 BP; 252 A; 451 C; 469 G; 298 T; 0 other;

  Query Match   7.6%; Score 32.2; DB 22; Length 1470;
  Best Local Similarity   50.3%; Pred. No. 7.2;
  Matches   79; Conservative   0; Mismatches   78; Indels   0; Gaps   0;

Qy   50 tttcaagttcaagcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaa 109
           |||||| ||||||||| | | | | |||| || | || ||| |
Db   75 tttcaaactcaagcaacacggcagcaccgtcagaaccgaaatgatcgctggggtgaccac 134

Qy   110 gttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgccca 169
              ||| || |||| |||| | |||| | ||| || | | || |
Db   135 cttcatcaccatggcctacatcatcttcgtcaaccccaacatcatggccgacgccggcat 194

Qy   170 ggaccatttcttcgacgtaatcacaaactctgtcggc 206
           |||||| || | || | || ||| |
Db   195 cgaccatggtgccgcttttgtcgccacctgcatcgcc 231

RESULT 14
AAZ32025
ID   AAZ32025 standard; DNA; 9810 BP.
XX
AC   AAZ32025;
XX
DT   10-JAN-2000   (first entry)
XX
DE   Human METH1 related EST AF018073.
XX
KW   Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW   cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW   angiogenesis inhibitor; abnormal wound healing; inflammation;
KW   rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW   diabetic retinopathy; macula degeneration; haemangioma; detection;
KW   arterial-venous malformation; immune deficiency; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937660-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-US01313.
XX
PR   23-JAN-1998;   98US-0072298.
PR   28-AUG-1998;   98US-0098539.
XX

CC   AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC   metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC   respectively. METH1 and METH2 have been found to be potent inhibitors of
CC   angiogenesis both in vitro and in vivo. They can be used for treating
CC   cancer and other disorders related to angiogenesis including abnormal
CC   wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC   endometrial bleeding disorders, diabetic retinopathy, some forms of
CC   macula degeneration, haemangiomas, and arterial-venous malformations.
CC   They may be useful in treating deficiencies or disorders of the immune
CC   system, by activating or inhibiting the proliferation, differentiation,
CC   or mobilisation (chemotaxis) of immune cells. The etiology of these
CC   immune deficiencies or disorders may be genetic, somatic, such as
CC   cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC   toxins), or infectious. They can also be used to treat inflammatory
CC   conditions, both chronic and acute conditions. The products can also be
CC   used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC   AAY49511 represent sequences given in the exemplification of the present
CC . invention.
XX
SQ   Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

      Query Match   7.6%;   Score 32.2; DB 20;   Length 9810;
      Best Local Similarity   43.3%;   Pred. No. 15;
      Matches 151; Conservative   0; Mismatches 198; Indels   0; Gaps   0;

Qy   78 gtcattagcatgtcctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacg 137
           |||·||··|||·||·····||·|·|·||··|·||···|||·||||||·||·······|
Db   3387 gtccttctcatcaccgtggtgggcggggtcctgctcgcgctcctgctcgaccagcccttc 3446

Qy   138 ttcaagaatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaac 197
           |····|·|·|·||||||·||·····|·····|·|·|·||||||·|·||···|||····|
Db   3447 tgggggcagggcatcgtgcgcgtgctggtgatcgctcccttcttcgtcatgcccaccgtc 3506

Qy   198 tctgtcggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgtt 257
           ||·|·····|·····|·····||·||····||·····||·||·|·|·|·||·|·|
Db   3507 tcggcgctggtctggaagaacatgttcatgaaccccgtgaacgggatgttcgcccatatc 3566

Qy   258 ggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaac 317
           |·|·|··||·|·····|·······|·····||·|·|··||·||···||·|
Db   3567 gcccgcgggctcggccttccgccgttcgacttcctgtcgcaggcgccgctggcctcgatc 3626

Qy   318 gtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttg 377

```
           |  ||||    ||     |  ||| |  |||      |      ||||      || |||     |
Db    3627 atcggcatcgtggcctggcagtggctgcccttcgccacgctgatccttctgacggcgctc 3686

Qy     378 atctggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
              |  |  | |||  ||| |||| ||      ||| | ||   | ||||
Db    3687 cagtcgctcgaccgcgagcagatggaggcggccgagatggacggcgcct 3735


RESULT  15
AAC90082
ID    AAC90082 standard; DNA; 9810 BP.
XX
AC    AAC90082;
XX
DT    19-MAR-2001  (first entry)
XX
DE    AF018073 cDNA clone.
XX
KW    METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW    cancer therapy; benign tumour; ocular angiogenic disease;
KW    rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW    vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW    scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW    coronary collateral; cerebral collateral; arteriovenous malformation;
KW    ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW    plaque neovascularisation; telangiectasia; haemophiliac joint; EST;
KW    angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW    Crohn's disease; atherosclerosis; birth control; ss.
XX
OS    Unidentified.
XX
PN    WO200071577-A1.
XX
PD    30-NOV-2000.
XX
PF    25-MAY-2000; 2000WO-US14462.
XX
PR    25-MAY-1999;   99US-0318208.
PR    20-JUL-1999;   99US-0144882.
PR    10-AUG-1999;   99US-0147823.
PR    13-AUG-1999;   99US-0373658.
PR    22-DEC-1999;   99US-0171503.
PR    22-FEB-2000; 2000US-0183792.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
PA    (SMIK ) SMITHKLINE BEECHAM CORP.
PA    (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA    (IRUE/) IRUELA-ARISPE L.
PA    (HAST/) HASTINGS G A.
PA    (RUBE/) RUBEN S M.
PA    (JONA/) JONAK Z L.
PA    (TRUL/) TRULLI S H.
PA    (FORN/) FORNWALD J A.
PA    (TERR/) TERRETT J A.
XX
PI    Iruela-Arispe L,  Hastings GA,  Ruben SM,  Jonak ZL,  Trulli SH;
PI    Fornwald JA,  Terrett JA;
```

```
XX
DR    WPI; 2001-025136/03.
XX
PT    METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT    inhibit angiogensis in the treatment of disorders such as cancer,
PT    rheumatoid arthritis and psoriasis -
XX
PS    Claim 7; Pages 653-659; 768pp; English.
XX
CC    The present invention relates to human METH1 and METH2, (ME for
CC    metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC    The present sequence is an expressed sequence tag (EST) for METH. METH
CC    can be used for inhibiting angiogenesis in an individual, and for
CC    treating cancer, benign tumours, an ocular angiogenic disease,
CC    rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC    vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC    scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC    coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC    ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC    neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC    fibromuscular dysplasia, wound granulation, Crohn's disease or
CC    atherosclerosis. METH can also be used in birth control. METH can also
CC    be used in diagnostic methods for the prognosis of cancer.
XX
SQ    Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;


   Query Match              7.6%;   Score 32.2;  DB 22;   Length 9810;
   Best Local Similarity   43.3%;   Pred. No. 15;
   Matches  151;  Conservative   0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy      78 gtcattagcatgtcctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacg 137
           ||| || ||| ||   || |  ||  | ||   | ||   ||| ||||| ||     |
Db    3387 gtccttctcatcaccgtggtgggcggggtcctgctcgcgctcctgctcgaccagcccttc 3446

Qy     138 ttcaagaatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaac 197
           |    | | |  ||||||| | |       |   | | ||||||| | |   |||    |
Db    3447 tgggggcagggcatcgtgcgcgtgctggtgatcgctcccttcttcgtcatgcccaccgtc 3506

Qy     198 tctgtcggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgtt 257
           || |     |      |     || ||      ||    || || | | | || | |
Db    3507 tcggcgctggtctggaagaacatgttcatgaacccgtgaacgggatgttcgcccatatc 3566

Qy     258 ggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaac 317
           | | |  || |  |        |        |    || || |  || ||    ||| |
Db    3567 gcccgcgggctcggccttccgccgttcgacttcctgtcgcaggcgccgctggcctcgatc 3626

Qy     318 gtaggcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttg 377
            | |||| || |   | ||| | ||| |    ||||    || ||| | |
Db    3627 atcggcatcgtggcctggcagtggctgcccttcgccacgctgatccttctgacggcgctc 3686

Qy     378 atctggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
            | | | ||| ||| |||| || ||    ||| | ||   | ||||
Db    3687 cagtcgctcgaccgcgagcagatggaggcggccgagatggacggcgcct 3735 .
```

Search completed: February 7, 2002, 11:01:03
Job time: 5049 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:22:27 ; Search time 172.96 Seconds
                                                   (without alignments)
                                                   557.815 Million cell updates/sec

Title: US-09-394-745-7565
Perfect score: 426
Sequence: 1 gggccgacccacgcgtccag..........catcgacacggtgcgagcct 426

Scoring table: IDENTITY_NUC
                     Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                     1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                     2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                     3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                     4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                     5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                     6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result   Query
   No.   Score Match Length DB   ID   Description
----
      1   37.6   8.8   1929 4   US-09-380-420C-1   Sequence 1, Appli
      2   36   8.5   2682 1   US-07-855-793-3   Sequence 3, Appli
      3   32.8   7.7 4403765 4   US-09-103-840A-2   Sequence 2, Appli
      4   32.8   7.7 4411529 4   US-09-103-840A-1   Sequence 1, Appli
      5   32.4   7.6   1032 4   US-09-193-503B-2   Sequence 2, Appli
      6   32.4   7.6   2346 4   US-09-193-503B-5   Sequence 5, Appli
  c   7   31.8   7.5   735 3   US-09-003-287-7   Sequence 7, Appli
      8   31.8   7.5 33529 4   US-09-144-085-3   Sequence 3, Appli
  c   9   31.4   7.4   1352 2   US-08-937-972-4   Sequence 4, Appli

```
    10   31.4   7.4    50341   1   US-08-247-901C-1        Sequence 1, Appli
    11   31.4   7.4    50341   2   US-09-075-904-1         Sequence 1, Appli
    12   31.4   7.4    52297   4   US-09-426-436-1         Sequence 1, Appli
    13   31.4   7.4    52297   4   US-08-705-557-1         Sequence 1, Appli
    14   31.2   7.3      804   4   US-08-998-416-881       Sequence 881, App
    15   31     7.3     4112   1   US-08-340-203A-2        Sequence 2, Appli
    16   31     7.3     4112   2   US-08-452-567-2         Sequence 2, Appli
    17   31     7.3     4112   2   US-08-452-427-2         Sequence 2, Appli
    18   31     7.3     4112   3   US-09-085-407-2         Sequence 2, Appli
    19   31     7.3     4616   1   US-08-340-203A-1        Sequence 1, Appli
    20   31     7.3     4616   2   US-08-452-567-1         Sequence 1, Appli
    21   31     7.3     4616   2   US-08-452-427-1         Sequence 1, Appli
    22   31     7.3     4616   3   US-09-085-407-1         Sequence 1, Appli
c   23   30.6   7.2     1730   6   5223391-8               Patent No. 5223391
c   24   30.6   7.2  4403765   4   US-09-103-840A-2        Sequence 2, Appli
    25   30.4   7.1    43280   2   US-08-804-227C-1        Sequence 1, Appli
c   26   30.4   7.1  4411529   4   US-09-103-840A-1        Sequence 1, Appli
c   27   30.2   7.1     1315   2   US-08-578-592-4         Sequence 4, Appli
c   28   30.2   7.1     1315   3   US-09-185-111-4         Sequence 4, Appli
    29   30.2   7.1     2742   4   US-09-232-468A-1        Sequence 1, Appli
    30   29.8   7.0     2159   3   US-08-286-870A-7        Sequence 7, Appli
c   31   29.6   6.9     1910   4   US-09-593-711A-3        Sequence 3, Appli
c   32   29.6   6.9     1914   1   US-07-601-094-1         Sequence 1, Appli
c   33   29.6   6.9     1914   1   US-08-012-735-1         Sequence 1, Appli
    34   29.6   6.9     4451   3   US-08-717-294-42        Sequence 42, Appl
c   35   29.4   6.9     1329   3   US-08-360-758-1         Sequence 1, Appli
c   36   29.4   6.9     1389   1   US-08-458-023B-1        Sequence 1, Appli
c   37   29.4   6.9     1389   3   US-09-111-556A-1        Sequence 1, Appli
    38   29.4   6.9     1602   4   US-09-333-423-3         Sequence 3, Appli
    39   29.4   6.9     9960   3   US-08-822-586-46        Sequence 46, Appl
    40   29.2   6.9    12912   2   US-08-460-751-1         Sequence 1, Appli
    41   29.2   6.9    14060   3   US-08-658-136-4         Sequence 4, Appli
    42   29.2   6.9    31571   1   US-08-323-443B-1        Sequence 1, Appli
    43   29.2   6.9    53526   3   US-08-658-136-2         Sequence 2, Appli
    44   29.2   6.9    53577   3   US-08-658-136-1         Sequence 1, Appli
    45   29     6.8     1011   2   US-09-013-634-3         Sequence 3, Appli
```

ALIGNMENTS

RESULT   1
US-09-380-420C-1
; Sequence 1, Application US/09380420C
; Patent No. 6300544
;   GENERAL INFORMATION:
;   APPLICANT: Halkier, Barbara
;   Bak, Soren
;   Kahn, Rachel
;   Moller, Birger
;   TITLE OF INVENTION: Cytochrome P450 Monooxygenases
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Syngenta Patent Dept.
;   STREET: 3054 Cornwallis Road
;   CITY: RTP
;   STATE: NC

US-09-380-420C-1

    Query Match   8.8%;   Score 37.6; DB 4; Length 1929;
    Best Local Similarity   49.5%; Pred. No. 0.052;
    Matches   97; Conservative   0; Mismatches   99; Indels   0; Gaps   0;

Qy   140 caagaatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactc 199
              || | | ||| | || || |||| ||||| ||| | ||| | |
Db   779 CATGGACATGATGGCCAGCTTCTCCGCCGAGGACTTCTTCCCCAACGCCGCCGGCCGCCT 838

Qy   200 tgtcggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttgg 259
              | || || |||||| | ||||| ||| | | | || | | | |
Db   839 CGCCGACCGCCTCTCGGGCTTCCTCGCCCGCCGCGAGCGCATCTTCAACGAGCTCGACGT 898

Qy   260 cgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
              | | | || | | | | || | || | |||||| | |||| ||| ||||
Db   899 CTTCTTCGAGAAGGTCATCGACCAGCACATGGACCCGGCGCGCCCCGTGCCGGACAACGG 958

Qy   320 aggcacactgataggc 335
              ||| || | | |
Db   959 CGGCGACCTCGTCGAC 974

RESULT   2
US-07-855-793-3
; Sequence 3, Application US/07855793

; Patent No. 5217880
; GENERAL INFORMATION:
;   APPLICANT: Masanori MITTA et al.
;   TITLE OF INVENTION: L-FUCOSE DEHYDROGENASE GENE,
;   TITLE OF INVENTION: MICROORGANISM HAVING SAID GENE AND PRODUCTION OF L-
FUCOSE
;   TITLE OF INVENTION: DEHYDROGENASE BY THE USE OF SAID MICROORGANISM
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack
;   STREET: 805 Fifteenth Street, N.W., #700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: DisplayWrite
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/855,793
;   FILING DATE: 19920323
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-8850
;   TELEFAX:
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2682 Base Pairs
;   TYPE: NUCLEIC ACID
;   . STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
;   ORGANISM: Arthrobacter Oxidans
;   STRAIN:
;   INDIVIDUAL ISOLATE:
;   DEVELOPMENTAL STAGE:
;   HAPLOTYPE:
;   TISSUE TYPE:
;   CELL TYPE:
;   CELL LINE:
;   ORGANELLE:
;   IMMEDIATE SOURCE:
;   LIBRARY:

  Query Match   8.5%;   Score 36;   DB 1;   Length 2682;
  Best Local Similarity   47.0%;   Pred. No. 0.18;
  Matches 111; Conservative   0; Mismatches 125; Indels   0; Gaps   0;

Qy   37 aggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcatgtcctctg 96
           |||| || || ||| | || | | || ||| |||| | ||| |
Db   1133 AGGACACCGAGGGCTTCGACGTCCCGGACGACCTCATCCGGGTCCGCGACTACTCCCGCG 1192

Qy   97 ttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtga 156
           | ||| || | | || | || | | | | | ||||| |
Db   1193 ACGGGGTGCTGCGCTCCATCGAGGAAAGCCTGCAGCGGCTGGGGACCGACCGGATCGACA 1252

Qy   157 gggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcgg 216
           | |||| ||| |||| | | | || | | | | || | | |||
Db   1253 TCGTCTACATCCACGACCCTGACGACTACTGGACCGAGGCCGTGGAGGGCGCCGCCCCGG 1312

Qy   217 cgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatc 272
           ||||| ||| | ||| || || || || | ||||| ||| |
Db   1313 CGCTGTCCGCCCTGCGGGACGAAGGGGTCATCAGGGCCTGGGGCGCAGGCATGAAC 1368

RESULT   3
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00

   Query Match   7.7%;   Score 32.8; DB 4; Length 4403765;
   Best Local Similarity   56.5%;   Pred. No. 22;
   Matches   61; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

Qy   154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213
              ||||||| || | || | || | ||| | || ||| | | | || | || |
Db 533801 tgagggtctgccacctgccccgtaatgtcgctggtatggcaagcaccgacgccgcggccc 533860

Qy   214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcg 261
              | | ||||| | | || || ||| ||| || | ||| | ||
Db 533861 aagagttgctccgcgacgcgttcacccggttgatcgaacatgtcgacg 533908

RESULT   4
US-09-103-840A-1

   Query Match   7.7%;   Score 32.8; DB 4; Length 4411529;
   Best Local Similarity   56.5%;   Pred. No. 22;
   Matches   61; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

Qy   154 tgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtct 213

```
        ||||||| || |   || |   || |      ||| |   || ||| | | | || | | || |
Db 532359 tgagggtctgccacctgccccgtaatgtcgctggtatggcaagcaccgacgccgcggccc 532418


Qy    214 cggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcg 261
            | |  ||||| |   | |  || ||   ||| ||| || | ||| | ||
Db 532419 aagagttgctccgcgacgcgttcacccggttgatcgaacatgtcgacg 532466



RESULT   5
US-09-193-503B-2
; Sequence 2, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
;   APPLICANT:  Baszczynski, Christopher L.
;   APPLICANT:  Lyznik, Leszek A.
;   APPLICANT:  Gordon-Kamm, William J.
;   APPLICANT:  Guan, Xueni
;   APPLICANT:  Rao, Guru
;   APPLICANT:  Tagliani, Laura A.
;   TITLE OF INVENTION:  A No. 6262341el Method For The Integration Of Foreign
DNA Into
;   TITLE OF INVENTION:  Eukaryotic Genomes
;   FILE REFERENCE:  5718-66 (amended listing)
;   CURRENT APPLICATION NUMBER: US/09/193,503B
;   CURRENT FILING DATE:  1998-11-17
;   PRIOR APPLICATION NUMBER:  60/099,435
;   PRIOR FILING DATE:  1998-09-08
;   PRIOR APPLICATION NUMBER:  60/056,627
;   PRIOR FILING DATE:  1997-11-18
;   PRIOR APPLICATION NUMBER:  60/065,613
;   PRIOR FILING DATE:  1997-11-18
;   NUMBER OF SEQ ID NOS:  11
;   SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1032
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
;   OTHER INFORMATION: sequence encoding Cre protein from Bacteriophage
;   OTHER INFORMATION: P1, maize preferred codons (moCRE)
US-09-193-503B-2


  Query Match              7.6%;  Score 32.4;  DB 4;  Length 1032;
  Best Local Similarity   54.1%;  Pred. No. 1.5;
  Matches   66; Conservative   0; Mismatches   56; Indels    0; Gaps    0;

Qy    299 gcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacct 358
            ||  ||||| |   ||||| |     | ||  .|| |   | || || | |   || |
Db     12 gctcacggttcaccagaaccttccggctcttccagtggacgcgacgtccgatgaagtcag 71

Qy    359 gacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggt 418
            || ||| |||| | || || || |  || |   ||| | | | || ||||||| |    | | |
Db     72 gaagaacctcatggacatgttccgcgacaggcaagcgttcagcgagcacacctggaagat 131

Qy    419 gc 420
```

```
        ||
Db    132 gc 133


RESULT    6
US-09-193-503B-5
; Sequence 5, Application US/09193503B
; Patent No. 6262341
; GENERAL INFORMATION:
;   APPLICANT:  Baszczynski, Christopher L.
;   APPLICANT:  Lyznik, Leszek A.
;   APPLICANT:  Gordon-Kamm, William J.
;   APPLICANT:  Guan, Xueni
;   APPLICANT:  Rao, Guru
;   APPLICANT:  Tagliani, Laura A.
;   TITLE OF INVENTION:  A No. 6262341el Method For The Integration Of Foreign
DNA Into
;   TITLE OF INVENTION:  Eukaryotic Genomes
;   FILE REFERENCE:  5718-66 (amended listing)
;   CURRENT APPLICATION NUMBER: US/09/193,503B
;   CURRENT FILING DATE:  1998-11-17
;   PRIOR APPLICATION NUMBER:  60/099,435
;   PRIOR FILING DATE:  1998-09-08
;   PRIOR APPLICATION NUMBER:  60/056,627
;   PRIOR FILING DATE:  1997-11-18
;   PRIOR APPLICATION NUMBER:  60/065,613
;   PRIOR FILING DATE:  1997-11-18
;   NUMBER OF SEQ ID NOS:  11
;   SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 2346
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: sequence
;   OTHER INFORMATION: encoding moCre:FLPm, Cre from Bacteriophage P1 and
;   OTHER INFORMATION: FLP from Saccharomyces, both maize preferred
;   OTHER INFORMATION: codons
;   NAME/KEY: CDS
;   LOCATION: (1)..(2346)
US-09-193-503B-5


  Query Match              7.6%;   Score 32.4;  DB 4;   Length 2346;
  Best Local Similarity   54.1%;   Pred. No. 2;
  Matches   66;  Conservative   0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    299 gcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacct 358
              ||  ||||| |    |||||| |      | || |  || |   | || || | | || |
Db     12 gctcacggttcaccagaaccttccggctcttccagtggacgcgacgtccgatgaagtcag 71

Qy    359 gacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggt 418
              || ||| |||| | || || || |  || |||| | |   | | |||||| |   | | |
Db     72 gaagaacctcatggacatgttccgcgacaggcaagcgttcagcgagcacacctggaagat 131


Qy    419 gc 420
              ||
```

RESULT   7
US-09-003-287-7/c
; Sequence 7, Application US/09003287
; Patent No. 6096947
; GENERAL INFORMATION:
; APPLICANT: Jayne, Susan
; APPLICANT: Barbour, Eric
; APPLICANT: Meyer, Terry
; TITLE OF INVENTION: METHODS FOR IMPROVING TRANSFORMATION EFFICIENCY
; FILE REFERENCE: moPAT_moCAH
; CURRENT APPLICATION NUMBER: US/09/003,287
; CURRENT FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 735
;   TYPE: DNA
;   ORGANISM: Myrothecium verrucaria
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (01)..(732)
US-09-003-287-7

  Query Match   7.5%;   Score 31.8; DB 3; Length 735;
  Best Local Similarity   53.7%;   Pred. No. 2;
  Matches   66; Conservative   0; Mismatches   57; Indels   0; Gaps   0;

Qy   262 ccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtag 321
          ||| | | || || |||| || ||| |!||| ||| || | ||| || |
Db   562 CCACCCAGGAGCCGAAGTCGTCGATGCCGTCGTAGGCGCCCACGTTGTCGTAGAGGGTGG 503

Qy   322 gcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatct 381
          | ||| || | || | || | || | || | | ||| || |||| || |
Db   502 CGAGCTGGATGAGCTGGCCGAGGAAGGTGATGTTGCCGTCGACGCCGACGTCCTCGTGGC 443

Qy   382 gga 384
          |||
Db   442 GGA 440

RESULT   8
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20

```
   Query Match                7.5%;   Score 31.8;  DB 4;   Length 33529;
   Best Local Similarity   52.7%;  Pred. No. 8.6;
   Matches  69;  Conservative   0;  Mismatches  62;  Indels   0;  Gaps    0;


Qy    291 acgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggca 350
          |||| ||||| || ||||| |   || ||   ||| ||| ||   |      | |
Db  27559 acgtacgcgcggccgcagctggcggtggtgagcggcgtgacgggcgagctcggtggcgaa 27618


Qy    351 gagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatc 410
          ||     ||||| |  | | | |||||| | | ||| |     | |||| | | |    | |
Db  27619 gaagcgctgatgtcggccgagtactgggtgaggcaggtgcgcgaggcggtgcgcttcctg 27678


Qy    411 gacacggtgcg 421
          ||| | ||||
Db  27679 gacgggatgcg 27689
```


RESULT   9
US-08-937-972-4/c

  Query Match   7.4%;   Score 31.4;   DB 2;   Length 1352;
  Best Local Similarity   54.9%;   Pred. No. 3.3;
  Matches   62; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

Qy   203 cggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgc 262
          || | || | | ||||||||| | | ||| || | ||||||| || | || |
Db   702 CGTCAGGGCCGAGCCGCTGCTCTTGGCCAGGTGCACCTTGTGGATGAGGCCAGGCAGCAC 643

Qy   263 catactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggaga 315
             | | || ||| | | | || | | || ||| || ||| ||
Db   642 GTGGCCATCCTCGATGTTGAAGTTGTCGTGGGGCCCGAAGACGTTGGTGGGGA 590

RESULT 10
US-08-247-901C-1
; Sequence 1, Application US/08247901C
; Patent No. 5750384
; GENERAL INFORMATION:
;   APPLICANT: Jacobs et al
;   TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amster, Rothstein & Ebenstein
;   STREET: 90 Park Avenue
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10016
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Processor (ASCII)
;   CURRENT APPLICATION DATA:

```
;       APPLICATION NUMBER:  US/08/247,901C
;       FILING DATE:  May 23, 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/057,531
;       FILING DATE:  April 29, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bogosian, Elizabeth A
;       REGISTRATION NUMBER:  39,911
;       REFERENCE/DOCKET NUMBER:  96700/273
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 697-5995
;       TELEFAX:  (212) 286-0854 or 286-0082
;       TELEX:  TWX 710-581-4766
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  50341
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  L5 shuttle phasmid sequence
;     HYPOTHETICAL:  No
;     ANTI-SENSE:
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;       ORGANISM:  L5 mycobacteriophage
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:
;       HAPLOTYPE:
;       TISSUE TYPE:
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:  No. 5750384e
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1
```

```
   Query Match              7.4%;   Score 31.4;   DB 1;   Length 50341;
   Best Local Similarity   49.7%;   Pred. No. 13;
   Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;


Qy    159 gcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcg 218
          || ||| || || |     ||      | |   || | | | |     |     |  | ||
Db  34426 GCGTACTCCGAGAAGATGTTGGCGACCTTCTGCAGCATCACAGCGAACGGCAGCGGGCCG 34485


Qy    219 ctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttg 278
          ||| | ||  | | ||  | | |  ||| ||||   ||| || | | | || ||| |
Db  34486 CTGGCCACTCCACCGAACGTCTTGAGCTTGGCCCCTTGCGGCCGGATGCGGCTCACGTCG 34545


Qy    279 tacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
          |||||    |  |  |||| |     ||| | | ||  |||
Db  34546 TACACCCGCTGGTAGTGGACCGTGCCGGGTCGGTAGTGCGT 34586



RESULT  11
US-09-075-904-1
; Sequence 1, Application US/09075904
; Patent No. 5994137
;   GENERAL INFORMATION:
;     APPLICANT:  Jacobs, et al.
;     TITLE OF INVENTION:  L5 SHUTTLE PHASMIDS
;     NUMBER OF SEQUENCES:  1
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Amster, Rothstein & Ebenstein
;       STREET:  90 Park Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10016
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Word Processor (ASCII)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/075,904
;       FILING DATE:  May 11, 1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/247,901
;       FILING DATE:  May 23, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bogosian, Elizabeth A
;       REGISTRATION NUMBER:  39,911
;       REFERENCE/DOCKET NUMBER:  96700/475
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 697-5995
;       TELEFAX:  (212) 286-0854 or 286-0082
;       TELEX:  TWX 710-581-4766
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  50341
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
```

```
;       TOPOLOGY:  linear
;    MOLECULE TYPE:
;       DESCRIPTION:  L5 shuttle phasmid sequence
;    HYPOTHETICAL:  No
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;       ORGANISM:  L5 mycobacteriophage
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:
;       HAPLOTYPE:
;       TISSUE TYPE:
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;    FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;    PUBLICATION INFORMATION:  No. 5994137e
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1
```

```
  Query Match              7.4%;   Score 31.4;  DB 2;  Length 50341;
  Best Local Similarity   49.7%;   Pred. No. 13;
  Matches   80;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     159 gcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcg 218
           || ||| || || |      ||      | |   || | | | |    |     | | ||
Db   34426 GCGTACTCCGAGAAGATGTTGGCGACCTTCTGCAGCATCACAGCGAACGGCAGCGGGCCG 34485

Qy     219 ctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttg 278
           ||| | ||  | | ||  | | |  ||| ·||||   |||    || | | || ||| |
Db   34486 CTGGCCACTCCACCGAACGTCTTGAGCTTGGCCCCTTGCGGCCGGATGCGGCTCACGTCG 34545

Qy     279 tacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
           ||||| |   | |  |||| |     ||| | | || |||
Db   34546 TACACCCGCTGGTAGTGGACCGTGCCGGGTCGGTAGTGCGT 34586
```

RESULT 12
US-09-426-436-1

; Sequence 1, Application US/09426436
; Patent No. 6225066
; GENERAL INFORMATION:
;   APPLICANT: William R. Jacobs, Jr.
;   APPLICANT: Barry R. Bloom
;   APPLICANT: Graham F. Hatfull
;   TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
;   TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amster, Rothstein & Ebenstein
;   STREET: 90 Park Avenue
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10016
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Processor (ASCII)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/426,436
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/705,557
;   FILING DATE:
;   APPLICATION NUMBER: US/08/057,531
;   FILING DATE:
;   APPLICATION NUMBER: 07/833,431
;   FILING DATE: February 7, 1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pasqualini, Patricia A.
;   REGISTRATION NUMBER: 34,894
;   REFERENCE/DOCKET NUMBER: 96700/238
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 697-5995
;   TELEFAX: (212) 286-0854 or 286-0082
;   TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 52297
;   TYPE: nucleotide
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE:
;   DESCRIPTION: phage genome sequence
;   HYPOTHETICAL: no
;   ANTI-SENSE: no
;   FRAGMENT TYPE: not applicable.
;   ORIGINAL SOURCE:
;   ORGANISM: mycobacteriophage L5
;   STRAIN: not applicable
;   INDIVIDUAL ISOLATE: L5
;   DEVELOPMENTAL STAGE: not applicable
;   HAPLOTYPE: not applicable

US-09-426-436-1

   Query Match   7.4%;   Score 31.4;   DB 4;   Length 52297;
   Best Local Similarity   49.7%;   Pred. No. 13;
  Matches   80; Conservative   0; Mismatches   81; Indels   0; Gaps   0;

Qy   159 gcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcg 218
          || ||| || || | || | | || | | | | | | | ||
Db 34323 GCGTACTCCGAGAAGATGTTGGCGACCTTCTGCAGCATCACAGCGAACGGCAGCGGGCCG 34382

Qy   219 ctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttg 278
          ||| | || | | || | | || | | | ||| |||| ||| || | | || ||| |
Db 34383 CTGGCCACTCCACCGAACGTCTTGAGCTTGGCCCCTTGCGGCCGGATGCGGCTCACGTCG 34442

Qy   279 tacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
          ||||| | | ||||| | ||| | | || |||
Db 34443 TACACCCGCTGGTAGTGGACCGTGCCGGGTCGGTAGTGCGT 34483

RESULT 13
US-08-705-557-1
; Sequence 1, Application US/08705557
; Patent No. 6300061
;   GENERAL INFORMATION:
;   APPLICANT: William R. Jacobs, Jr.
;   APPLICANT: Barry R. Bloom
;   APPLICANT: Graham F. Hatfull
;   TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
;   TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amster, Rothstein & Ebenstein
;   STREET: 90 Park Avenue
;   CITY: New York
;   STATE: New York

```
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,557
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/057,531
;      FILING DATE:
;      APPLICATION NUMBER:  07/833,431
;      FILING DATE:  February 7, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pasqualini, Patricia A.
;      REGISTRATION NUMBER:  34,894
;      REFERENCE/DOCKET NUMBER:  96700/238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52297
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  phage genome sequence
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:  not applicable.
;    ORIGINAL SOURCE:
;      ORGANISM:  mycobacteriophage L5
;      STRAIN:  not applicable
;      INDIVIDUAL ISOLATE:  L5
;      DEVELOPMENTAL STAGE:  not applicable
;      HAPLOTYPE:  not applicable
;      TISSUE TYPE:  not applicable
;      CELL TYPE:  not applicable
;      CELL LINE:  not applicable
;      ORGANELLE:  not applicable
;    IMMEDIATE SOURCE:  mycobacteriophage L5 particles
;    POSITION IN GENOME:  entire genome
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Hatfull and Sarkis
;      TITLE:  DNA Sequence, Structure and Gene
;      TITLE:  Expression of Mycobacteriophage L5:
;      TITLE:  A Phage System for Mycobacterial
```

   Query Match   7.4%;   Score 31.4;   DB 4;   Length 52297;
   Best Local Similarity   49.7%;   Pred. No. 13;
   Matches   80; Conservative   0; Mismatches   81; Indels   0; Gaps   0;

Qy   159 gcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcg 218
           || ||| || || | || | | || | | | | | | | || |
Db   34323 GCGTACTCCGAGAAGATGTTGGCGACCTTCTGCAGCATCACAGCGAACGGCAGCGGGCCG 34382

Qy   219 ctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttg 278
           ||| | || | | || | | | ||| |||| ||| || | | || ||| |
Db   34383 CTGGCCACTCCACCGAACGTCTTGAGCTTGGCCCCTTGCGGCCGGATGCGGCTCACGTCG 34442

Qy   279 tacacgatcacgacgtgggcgcgaacggtgctggagaacgt 319
           ||||| | | |||| | ||| | | || |||
Db   34443 TACACCCGCTGGTAGTGGACCGTGCCGGGTCGGTAGTGCGT 34483

RESULT 14
US-08-998-416-881
; Sequence 881, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
;   APPLICANT: Philippsen, Peter
;   APPLICANT: Pohlmann, Rainer
;   APPLICANT: Steiner, Sabine
;   APPLICANT: Mohr, Christine
;   APPLICANT: Wendland, Jurgen
;   APPLICANT: Knechtle, Philipp
;   APPLICANT: Rebischung, Corinne
;   TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;   TITLE OF INVENTION: AND USES THEREOF
;   NUMBER OF SEQUENCES: 1152
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: No. 6239264artis Corporation
;   STREET: 3054 Cornwallis Road
;   CITY: Research Triangle Park
;   STATE: No. 6239264th Carolina
;   COUNTRY: USA
;   ZIP: 27709
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/998,416
;   FILING DATE: 24-DEC-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:

   Query Match   7.3%;   Score 31.2;   DB 4;   Length 804;
   Best Local Similarity   46.7%;   Pred. No. 3.1;
   Matches   99; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

Qy   5 cgacccacgcgtccagatgagatgacccaatcaggagcacgcggatttcaagttcaagca 64
            | ||| | |||| | || | | ||| || | | ||.|| ||||
Db   16 CCACCACAACTTCCACGTCCACTGCATCTACCAGTGGCTCAACACCTCCACGTCCAAGGG 75

Qy   65 agagctctggatggtcattagcatgtcctctgttgcggtcgtgäagttcttcctcatgct 124
             |||| || || || | || | | | | | ||
Db   76 CCTCTGTCCGATGTGCAGGCAAGCGTTTTCACTCCGGGAGGGCATCCGCATTAACGAGCC 135

Qy   125 ctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttcttcga 184
            | || || | |||| |||| | | | ||| ||| | ||| | | ||
Db   136 CCACCGCGACAAGTTCGAGAAGGTGTTGATGAAGGCGCGCCAGCAGAGCGTGGTGAGCGT 195

Qy   185 cgtaatcacaaactctgtcggcctggtctcgg 216   ·
            || | | ||| | ||||| | || | ||
Db   196 CGCGGGCGCCAACCCGGTCGGGCCGGACCAGG 227

RESULT 15
US-08-340-203A-2
; Sequence 2, Application US/08340203A
; Patent No. 5756668
; GENERAL INFORMATION:
;   APPLICANT: Baylin, Stephen B.
;   APPLICANT: Wales, Michele M.
;   TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 4225 Executive Square, Suite 1400
;   CITY: La Jolla
;   STATE: California
;   COUNTRY: USA

;   ZIP: 92037
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/340,203A
;   FILING DATE: 15-NOV-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Haile, Ph.D., Lisa A.
;   REGISTRATION NUMBER: P-38,347
;   REFERENCE/DOCKET NUMBER: 07265/039001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 455-5100
;   TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4112 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   IMMEDIATE SOURCE:
;   CLONE: HIC-1 coding polynucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1086..2726
US-08-340-203A-2

  Query Match   7.3%; Score 31; DB 1; Length 4112;
  Best Local Similarity 59.8%; Pred. No. 6.6;
  Matches   52; Conservative   0; Mismatches   35; Indels   0; Gaps   0;

Qy   333 ggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatctggaaccaccat 392
          | |||| |||| || | ||| ||| || || |||| |||| | ||| | |
Db   2524 GACAAGGCGGCCGCGACCGAGCTGCTGGCGCAGACCACGCACTTCCTGCACGACCCCAAG 2583

Qy   393 gaggagatccagcacatcgacacggtg 419
          | || | | || | || || || ||
Db   2584 GTGGCGCTGGAGAGCCTCTACCCGCTG 2610

Search completed: February 7, 2002, 11:42:57
Job time: 9143 sec

                        GenCore version 4.5
                Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 08:21:01 ; Search time 4942.22 Seconds
                                     (without alignments)

```
Title:            US-09-394-745-7565
Perfect score:    426
Sequence:         1 gggccgacccacgcgtccag..........catcgacacggtgcgagcct 426


Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0


Searched:         11351937 seqs, 5372889281 residues


Total number of hits satisfying chosen parameters:      22703874


Minimum DB seq length: 0
Maximum DB seq length: 2000000000


Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries


Database :        EST:*
                  1:  em_estfun:*
                  2:  em_esthum:*
                  3:  em_estin:*
                  4:  em_estom:*
                  5:  em_estpl:*
                  6:  em_estba:*
                  7:  em_estro:*
                  8:  em_estov:*
                  9:  em_htc:*
                  10: gb_est1:*
                  11: gb_est2:*
                  12: gb_htc:*
                  13: gb_gss:*
                  14: em_gss_fun:*
                  15: em_gss_hum:*
                  16: em_gss_inv:*
                  17: em_gss_pln:*
                  18: em_gss_pro:*
                  19: em_gss_rod:*
                  20: em_gss_vrt:*
                  21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	158	37.1	689	10	AI726300	AI726300 BNLGHi554
2	152	35.7	497	10	AW458679	AW458679 sh12c08.y
3	151.6	35.6	622	10	BE034614	BE034614 ML04B02 M
4	151.6	35.6	622	10	BE034615	BE034615 ML04B03 M
5	151.6	35.6	625	10	BE033763	BE033763 MF06B02 M

```
     6   149.4   35.1    495   10   AI966737        AI966737 sc56g04.y
 c   7   148.6   34.9    699   11   BE821231        BE821231 GM700024A
     8   146.8   34.5    674   11   BI263615        BI263615 NF090C09P
 c   9   141.4   33.2    674   11   BI422631        BI422631 EST533297
    10   140.4   33.0    619   10   AW396729        AW396729 sg80a05.y
    11   125.2   29.4    507   10   AW756123        AW756123 sl16b11.y
    12   121.6   28.5    597   11   BG154726        BG154726 sab38c04.
    13   117.4   27.6    401   10   AU031216        AU031216 AU031216
    14   115.6   27.1    374   11   BG887449        BG887449 EST513300
    15   107.6   25.3    729   11   BG588773        BG588773 EST490582
    16   106.6   25.0    315   11   BI118775        BI118775 EST163 Di
    17    98.8   23.2    460   11   W43358          W43358 22735 Lambd
    18    95.8   22.5    317   11   BI118802        BI118802 EST190 Di
    19    90.8   21.3    583   10   BE366405        BE366405 PI1_32_A0
    20    89.4   21.0    366   11   D47505          D47505 RICS13048A
    21    89.4   21.0    620   10   AW399887        AW399887 707052E06
    22    87.2   20.5    301   10   AA067522        AA067522 26383 Lam
    23    83.8   19.7    476   11   BG136826        BG136826 EST477268
    24    81.8   19.2    336   11   D46323          D46323 RICS10925A
    25    78.4   18.4    584   11   BI424328        BI424328 saf31e07.
    26    78.4   18.4    615   11   BG933069        BG933069 WS1_3_G11
 c  27    77.4   18.2    503   10   AV542466        AV542466 AV542466
    28    75.2   17.7    655   11   BF274320        BF274320 GA__Eb002
 c  29    72.6   17.0    677   10   AW505883        AW505883 GE1117 Gl
 c  30    71.2   16.7    517   10   AW031194        AW031194 EST274732
    31    70      16.4    430   10   AA394424        AA394424 26007 Lam
 c  32    68.2   16.0    514   10   AI987336        AI987336 660004B03
    33    66.6   15.6    510   10   BE474290        BE474290 sp61e12.y
    34    62.8   14.7    511   11   BG650823        BG650823 sad93a12.
    35    55.2   13.0    530   10   AI100583        AI100583 34958 Lam
    36    51.2   12.0    400   11   T04741          T04741 788 Lambda-
    37    48.4   11.4    157   10   AW234908        AW234908 sf20e10.y
    38    47.8   11.2    236   10   BE130052        BE130052 945035G09
    39    47.4   11.1    438   11   H37731          H37731 15860 Lambd
 c  40    47.2   11.1    897   13   BH135244        BH135244 ENTON12TF
 c  41    47      11.0    836   13   AZ536146        AZ536146 ENTCN35TR
    42    47      11.0    841   13   AZ693266        AZ693266 ENTML12TF
    43    43.8   10.3    874   13   AZ549401        AZ549401 ENTES23TR
    44    43      10.1    782   11   BG647203        BG647203 EST508822
 c  45    42.8   10.0    551   13   B26749          B26749 T18O5TF TAM
```

ALIGNMENTS

RESULT   1
AI726300
LOCUS   AI726300   689 bp   mRNA   EST   11-JUN-1999
DEFINITION BNLGHi5540 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AC004218) unknown protein [Arabidopsis thaliana], mRNA
            sequence.
ACCESSION   AI726300
VERSION   AI726300.1 GI:5045152
KEYWORDS   EST.
SOURCE   upland cotton.
  ORGANISM Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                  Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 689)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE   ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Ben Burr
             Biology Department
             Brookhaven National Laboratory
             Upton, NY 11973, USA
             Tel: 516-344-3396
             Fax: 516-344-3407
             Email: burr@bnlux1.bnl.gov
             Seq primer: T3 Primer.
FEATURES   Location/Qualifiers
     source   1. .689
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT   193 a   140 c   154 g   200 t   2 others
ORIGIN

   Query Match   37.1%;   Score 158;   DB 10;   Length 689;
   Best Local Similarity   62.3%;   Pred. No. 5.5e-33;
   Matches 248; Conservative   0; Mismatches 150; Indels   0; Gaps   0;

Qy   29 acccaatcaggagcacgcggatttcaagttcaagcaagagctctggatggtcattagcat 88
            | | ||| |||| | | | | ||||||||| ||| | || | ||
Db   11 ATCAAATGAGGATGAATTCAACCTGACAAATGAGCAAGAGAGGTGGGTTGTTGGCATTAT 70

Qy   89 gtcctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaagaatga 148
            | || | | ||| | ||||||||| |||| ||||| || || | || ||
Db   71 GCTTGGAGTGACTCTGACAAAGCTTGTCCTCATGTTCTATTGCCGCACATTTACAAACGA 130

Qy   149 gatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtcggcct 208
            ||||| | · || || || ||||| || ||||| || || ||||||||||| | |||||
Db   131 AATCGTTAAAGCTTATGCTCAGGATCACTTCTTTGATGTTATCACAAACATCATTGGCCT 190

Qy   209 ggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgccatact 268
            || | | ||| || ||| | | ||||||||||||||||||| || || |
Db   191 TGTTGCTGTGCTACTTGCTAAGTACATCGACGATTGGATGGACCCTGTTGGAGCTATCAT 250

Qy   269 gatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtaggcacact 328
            | || ||||||||| || ||| ||| || || || | | ||||||||| | || |
Db   251 TCTGGCTTTGTACACAATACGGACATGGTCGATGACAGTATTAGAGAACGTGAACTCATT 310

Qy   329 gataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatctggaacca 388
            | | || | || || | |||| || || | ||| || || || || ||||||||
Db   311 GGTTGGAAGATCAGCAGCTCCAGAATATCTTCAGAAACTGACCTATCTGTGTTGGAACCA 370

Qy   389 ccatgaggagatccagcacatcgacacggtgcgagcct 426

```
           |||| ||| || || |||||||| ||||| ||||| |
Db    371 CCATAAGGCCATAAAGAACATCGATACGGTCCGAGCTT 408
```


RESULT   2
AW458679
LOCUS   AW458679   497 bp   mRNA   EST   17-JUL-2000
DEFINITION   sh12c08.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
             Gm-c1016-4551 5' similar to TR:O80632 O80632 F12L6.11 PROTEIN. ;,
             mRNA sequence.
ACCESSION   AW458679
VERSION   AW458679.1 GI:7028896
KEYWORDS   EST.
SOURCE   soybean.
  ORGANISM   Glycine max
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
             Glycine.
REFERENCE   1 (bases 1 to 497)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
             ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
             Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
             ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
             ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
             ,R., Waterston,R. and Wilson,R.
  TITLE   Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Shoemaker R/Public Soybean EST Project
             Public Soybean EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available through: Genome Systems, Inc. 4633 World
             Parkway Circle St. Louis, Missouri 63134 For further information
             call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
             427-3324 or contact: clones@genomesystems.com or
             info@genomesystems.com web site: www.genomesystems.com
             Insert Length: 812   Std Error: 0.00
             Seq primer: -40RP from Gibco
             High quality sequence stop: 419.
FEATURES   Location/Qualifiers
     source   1. .497
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-4551"
                     /clone_lib="Gm-c1016"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
```

> (dT) sequence with a XhoI restriction site. EcoRI adapters
> were ligated to the blunt-ended cDNA fragments followed by
> XhoI digestion. The cDNA fragments were directionally
> cloned into the EcoRI-XhoI restriction site of the
> pBluescript vector. The ligated cDNA fragments were
> transformed into XL10-Gold host cells. This library was
> constructed by Dr. Randy Shoemaker and Dr. John
> Erpelding."
BASE COUNT      147 a      102 c      103 g      145 t
ORIGIN


   Query Match              35.7%;   Score 152;  DB 10;   Length 497;
   Best Local Similarity   67.2%;   Pred. No. 2.2e-31;
   Matches 215;  Conservative    0;  Mismatches 105;  Indels    0;  Gaps    0;


Qy     107 gaagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgc 166
           || |||| | || ||| |||||||| |||| |||||| |||||| || || || || || ||
Db       5 GAGGTTCATTCTTATGGTCTACTGTCGAAGATTCAAAAATGAAATTGTTAGAGCATATGC 64

Qy     167 ccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgc 226
           || || || || || || || || || || || || ||||| || | |   | || |||| ||
Db      65 ACAAGATCACTTTTTTGATGTCATTACTAATTCTGTTGGATTAGCTGCTGCTGTGCTAGC 124

Qy     227 tgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgat 286
           .||||  || | | |||||||||| || ||    || || ||  | || || ||||| || ||
Db     125 TGTCAAGTTCTACTGGTGGATTGATCCAACAGGAGCTATTATTATAGCATTGTATACAAT 184

Qy     287 cacgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgcc 346
           || || |||||     || ||  | |||||| || |  |||| ||||| | | | || ||
Db     185 CAATACATGGGCCAAGACTGTCATTGAGAATGTTTGGTCACTCATAGGAAGGACAGCACC 244

Qy     347 ggcagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagca 406
            | || |  || | ||| | ||| | | | || ||||| |||||||| |||||| ||||
Db     245 ACCTGATTTTCTAGCCAAGTTAACTTTCCTCATATGGAATCACCATGAACAGATCAAGCA 304

Qy     407 catcgacacggtgcgagcct 426
           ||| || || || |||| |
Db     305 CATAGATACTGTTAGAGCAT 324


RESULT    3
BE034614
LOCUS       BE034614       622 bp     mRNA               EST         07-JUN-2000
DEFINITION  ML04B02 ML Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE034614
VERSION     BE034614.1  GI:8329623
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1 (bases 1 to 622)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.

      FEATURES                Location/Qualifiers
           source             1. .622
                              /organism="Mesembryanthemum crystallinum"
                              /db_xref="taxon:3544"
                              /clone_lib="ML"
                              /tissue_type="flowers and developing seedpods"
                              /dev_stage=">12 weeks"
                              /note="6 weeks in 500mM NaCL"
      BASE COUNT      145 a    144 c     146 g     186 t       1 others
      ORIGIN


      Query Match              35.6%;   Score 151.6;   DB 10;   Length 622;
      Best Local Similarity    63.4%;   Pred. No. 3e-31;
      Matches  232;  Conservative    0;   Mismatches  134;   Indels    0;   Gaps    0;

Qy      61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
            | ||||||    ||| | ||     |   ||| |||||||| | | || ||| |  |  |
Db     156 ACCAAGAGAGATGGCTTGTGGGCATTATGCTCTCTGTTACTCTGGTTAAGCTTCTATTGG 215

Qy     121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
            | || ||||||||| | |||| ||||||||| || | ||||||||| ||||||||| || |
Db     216 TCCTTTACTGCCGCTCCTTCACCAATGAGATAGTCAAAGCCTACGCGCAGGACCACTTTT 275

Qy     181 tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
             ·  | || || || || ||| |||    | ||||| |  | || || || |||        |
Db     276 TTGATGTTATTACCAACATCATTGGCCTCATTGCTGCTCTCCTGGCTAATTACGTTAGTG 335

Qy     241 ggtggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgc 300
            ||||||||| |||||||| || ||  |   | || || |||||  |||    ||||||| |
Db     336 ACTGGATGGATCCTGTTGGAGCTATCATTCTTGCTTTCTACACTATCCGAACGTGGTCAA 395

Qy     301 gaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctga 360
            || ||| |||| || |||     |  | | || || || ||   || |||| || |||
Db     396 TGACTGTGTTGGAAAATGTAAATTCGTTAGTTGGAAAATCTGCCACGCCAGACTATCTGC 455

Qy     361 cgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
            ||| || || || |      ||||||||||| |||   ||  |||||||||||| || |
Db     456 AGAAACTAACTTATCTTTGTTGGAACCACCACAAGGCTGTCAGGCACATCGACACAGTCC 515

Qy     421 gagcct 426
            | ||| |
Db     516 GCGCAT 521


RESULT     4
BE034615
LOCUS          BE034615      622 bp    mRNA              EST        07-JUN-2000

```
DEFINITION  ML04B03 ML Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE034615
VERSION     BE034615.1  GI:8329624
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            An open reading frame exists.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="ML"
                     /tissue_type="flowers and developing seedpods"
                     /dev_stage=">12 weeks"
                     /note="6 weeks in 500mM NaCL"
BASE COUNT      145 a     144 c     145 g     187 t        1 others
ORIGIN
```

```
     Query Match              35.6%;   Score 151.6;  DB 10;  Length 622;
     Best Local Similarity    63.4%;   Pred. No. 3e-31;
     Matches  232;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;


Qy      61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
            | ||||||   ||| | ||     |  ||| |||||||| | | || ||| |  |  |
Db     156 ACCAAGAGAGATGGCTTGTGGGCATTATGCTCTCTGTTACTCTGGTTAAGCTTCTATTGG 215


Qy     121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
            | || ||||||||| | |||| ||||||||| || | ||||||||| ||||||||| || |
Db     216 TCCTTTACTGCCGCTCCTTCACCAATGAGATAGTCAAAGCCTACGCGCAGGACCACTTTT 275


Qy     181 tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
            | || || || || ||| |||     | |||||  |  | || || || |||        |
Db     276 TTGATGTTATTACCAACATCATTGGCCTCATTGCTGCTCTCCTGGCTAATTACGTTAGTG 335


Qy     241 ggtggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgc 300
            ||||||||| |||||||| || ||   |  | || || ||||| |||     |||||| |
Db     336 ACTGGATGGATCCTGTTGGAGCTATCATTCTTGCTTTCTACACTATCCGAACGTGGTCAA 395


Qy     301 gaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctga 360
            || ||| |||| || ||| |    |   | || || || || || |||| |||| || |||
Db     396 TGACTGTGTTGGAAAATGTAAATTCGTTAGTTGGAAAATCTGCCACGCCAGACTATCTGC 455
```

```
Qy    361 cgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
          ||| || || ||  |     |||||||||||| |||   ||  |||||||||||| || |
Db    456 AGAAACTAACTTATCTTTGTTGGAACCACCACAAGGCTGTCAGGCACATCGACACAGTCC 515


Qy    421 gagcct 426
          |  || |
Db    516 GCGCAT 521



RESULT   5
BE033763
LOCUS       BE033763      625 bp    mRNA          EST       07-JUN-2000
DEFINITION  MF06B02 MF Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE033763
VERSION     BE033763.1  GI:8328772
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
    TITLE   Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            An open reading frame exists.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MF"
                     /tissue_type="Root"
                     /dev_stage="5-6 weeks old"
                     /note="Vector: Bluescript SK+; Site_1: EcorI; Site_2:
                     XhoI"
BASE COUNT      143 a     145 c     147 g     190 t
ORIGIN


  Query Match            35.6%;   Score 151.6;  DB 10;  Length 625;
  Best Local Similarity  63.4%;   Pred. No. 3e-31;
  Matches  232;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy     61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
          | |||||||   ||| | ||    |  |||  ||||||| | | || ||| |  |  |  |
Db    152 ACCAAGAGAGATGGCTTGTGGGCATTATGCTCTCTGTTACTCTGGTTAAGCTTCTATTGG 211


Qy    121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
```

```
            |  ||  ||||I'|||   |  ||||   ||||||||| || |   |||||||||| |||||||| || |
Db    212 TCCTTTACTGCCGCTCCTTCACCAATGAGATAGTCAAAGCCTACGCGCAGGACCACTTTT 271

Qy    181 tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
            |  ||  || || || |||      | |||||  |  |  || || || |||            |
Db    272 TTGATGTTATTACCAACATCATTGGCCTCATTGCTGCTCTCCTGGCTAATTACGTTAGTG 331

Qy    241 ggtggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgc 300
            ||||||||| |||||||| || ||  |  | || || ||||| |||      |||||| |
Db    332 ACTGGATGGATCCTGTTGGAGCTATCATTCTTGCTTTCTACACTATCCGAACGTGGTCAA 391

Qy    301 gaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctga 360
            ||  |||  ||||  ||  |||      |  |  |  | ||  || || || ||  ||||  || |||
Db    392 TGACTGTGTTGGAAAATGTAAATTCGTTAGTTGGAAAATCTGCCACGCCAGACTATCTGC 451

Qy    361 cgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
            ||| || || ||  |        |||||||||||  ·|||     || |||||||||||||| || |
Db    452 AGAAACTAACTTATCTTTGTTGGAACCACCACAAGGCTGTCAGGCACATCGACACAGTCC 511

Qy    421 gagcct 426
            |  || |
Db    512 GCGCAT 517
```

RESULT   6
AI966737
LOCUS   AI966737   495 bp   mRNA   EST   13-DEC-1999
DEFINITION sc56g04.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-463 5' similar to TR:O80632 O80632 F12L6.11 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AI966737
VERSION   AI966737.1 GI:5761378
KEYWORDS   EST.
SOURCE   soybean.
  ORGANISM Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 495)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE   Public Soybean EST Project
  JOURNAL   .Unpublished (1999)
  COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information

Possible reversed clone: similarity on wrong strand
High quality sequence stop: 392.

```
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-463"
                     /clone_lib="Gm-c1016"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT       147 a      104 c      104 g      140 t
ORIGIN
```

```
  Query Match              35.1%;  Score 149.4;  DB 10;  Length 495;
  Best Local Similarity    66.8%;  Pred. No. 1.1e-30;
  Matches  213;  Conservative   0;  Mismatches  106;  Indels    0;  Gaps    0;


Qy    108 aagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcc 167
              |  |||| | || ||| |||||||| ||||  ||||| ||||| || |  || || || ||
Db      4 AGGTTCATTCTTATGGTCTACTGTCGAAGATTCAAAAATGAAATTGGTAGAGCATATGCA 63

Qy    168 caggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgct 227
              || || || || |  || || || || || |||| || | |    | || |||| ||| 
Db     64 CAAGATCACTTTTCTGATGTCATTACTAATTCTGTTGGATTAGCTGCTGCTGTGCTAGCT 123

Qy    228 gtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgatc 287
              |||    |||| | ||||||||| || ||      || || || | || || ||||| || |||
Db    124 GTCAAGTACTACTGGTGGATTGATCCAACAGGAGCTATTATTATAGCATTGTATACAATC 183

Qy    288 acgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccg 347
              |  || |||||      || ||  | |||||| ||  |  |||| |||||  | |  | || ||
Db    184 AATACATGGGCCAAGACTGTCATTGAGAATGTTTGGTCACTCATAGGAAGGACAGCACCA 243

Qy    348 gcagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcac 407
              | ||| |  || | ||| | || | | | || ||||| ||||||||| |||||| |||||
Db    244 CCTGATTTTCTAGCCAAGTTAACTTTCCTCATATGGAATCACCATGAACAGATCAAGCAC 303

Qy    408 atcgacacggtgcgagcct 426
              || || || || |||| |
```

Db   304 ATAGATACTGTTAGAGCAT 322

RESULT   7
BE821231/c
LOCUS   BE821231   699 bp   mRNA   . EST   24-MAY-2001
DEFINITION   GM700024A10F6 Gm-r1070 Glycine max cDNA clone Gm-r1070-3707 3',
             mRNA sequence.
ACCESSION   BE821231
VERSION   BE821231.1 GI:10253465
KEYWORDS   EST.
SOURCE   soybean.
  ORGANISM   Glycine max
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
             Glycine.
REFERENCE   1   (bases 1 to 699)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
             Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
    TITLE   A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
  COMMENT   Other_ESTs: AI966737 corresponding to Gm-c1016-463 (5')
             Contact: Vodkin, L.O., PI, A Functional Genomics Program for
             Soybean (NSF 9872565)
             Lewin, H. A., Director, Keck Center for Comparative and Functional
             Genomics
             University of Illinois
             Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
             Tel: (217) 244-6147
             Fax: (217) 333-4582
             Email: l-vodkin@uiuc.edu
             This clone is available through: Genome Systems, Inc. 4633 World
             Parkway Circle St. Louis, Missouri 63134. For further information
             call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
             427-3324 or contact:clones@genomesystems.com or info@genome
             ystems.com web site:www.genomesystems.com
             Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES   Location/Qualifiers
     source   1. .699
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-3707"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216 clones selected from cDNA libraries from
                     various tissues and stages of development of soybean that
                     represent 2,639 sequences from immature cotyledons, 1,770
                     from immature seed coats, 3,938 from flowers, and 869
                     from young pods. The 5' ESTs of the source clones from
                     the different libraries was used to select singletons, or
                     a representative of each contig, which were reracked to
                     form library Gm-r1070. The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced at the 3' end. The
                     contig analysis to select unique genes was performed by
                     the laboratory of Ernest Retzel, Center for Computational
                     Genomics and Bioinformatics, University of Minnesota,

BASE COUNT   208 a   137 c   131 g   197 t   26 others
ORIGIN

   Query Match   34.9%;   Score 148.6; DB 11; Length 699;
   Best Local Similarity 66.1%;   Pred. No. 2.1e-30;
   Matches 211; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

Qy   108 aagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcc 167
          | ||||| | ||| ||||||| |||| ||||| ||||| || || || || || ||
Db   660 AGGTTCATNNNTATGGNCTACTGTCGAAGATTCAAAAATGAAATTGTTAGAGCATATGCA 601

Qy   168 caggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgct 227
          || || || || || || || || || || || |||||| || | | | || |||| |||
Db   600 CAAGATCACTTTTTTGATGTCATTACTAATTCTGTTGGATTAGCTGCTGCTGTGCTAGCT 541

Qy   228 gtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacgatc 287
          ||| || | | |||||||||| || || || || || | || || ||||| || |||
Db   540 GTCAAGTTCTACTGGTGGATTGATCCAACAGGAGCTATTATTATAGCATTGTATACAATC 481

Qy   288 acgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccg 347
          | || |||||| || || | |||||| || | |||| ||||| | | | || ||
Db   480 AATACATGGGCCAAGACTGTCATTGAGAATGTTTGGTCACTCATAGGAAGGACAGCACCA 421

Qy   348 gcagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcac 407
          | ||| | || | ||| | || | | | || |||| ||||| |||||||| ||||| |||||
Db   420 CCTGATTTTCTAGCCAAGTTAACTTTCCTCATATGGAATCACCATGAACAGATCAAGCAC 361

Qy   408 atcgacacggtgcgagcct 426
          || || || || ||||| |
Db   360 ATAGATACTGTTAGAGCAT 342

RESULT   8
BI263615
LOCUS   BI263615   674 bp   mRNA   EST   18-JUL-2001
DEFINITION NF090C09PL1F1070 Phosphate starved leaf Medicago truncatula cDNA
            clone NF090C09PL 5', mRNA sequence.
ACCESSION   BI263615
VERSION   BI263615.1 GI:14865019
KEYWORDS   EST.
SOURCE   barrel medic.
  ORGANISM Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
                    Medicago.
REFERENCE   1 (bases 1 to 674)
   AUTHORS   Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
             ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
     TITLE   Expressed Sequence Tags from the Samuel Roberts Noble Foundation
             Medicago truncatula phosphate-starved leaf library
   JOURNAL   Unpublished (2000)
   COMMENT   Contact: Harrison MJ
             Plant Biology Division
             The Samuel Roberts Noble Foundation
             2510 Sam Noble Parkway, Ardmore, OK 73402, USA
             Tel: 580 221 7325
             Fax: 580 221 7380
             Email: mjharrison@noble.org
             Insert Length: 674   Std Error: 0.00
             Plate: 090 row: C column: 09
             Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES   Location/Qualifiers
     source   1. .674
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF090C09PL"
                     /clone_lib="Phosphate starved leaf"
                     /tissue_type="leaf"
                     /dev_stage="trifoliate"
                     /note="Vector: Lambda Zap; At the trifoliate stage, M.
                     truncatula plants were transplanted to phosphate-free sand
                     and grown for a further 30 days. During this 30 day
                     period, the plants were fertilized twice weekly with 1/2
                     Hoaglands solution containing only 20uM potassium
                     phosphate. RNA was prepared from above ground tissues."
BASE COUNT   171 a   138 c   152 g   211 t   2 others
ORIGIN

   Query Match   34.5%;   Score 146.8; DB 11;   Length 674;
   Best Local Similarity 62.6%;   Pred. No. 6.3e-30;
   Matches 229; Conservative   0; Mismatches 137; Indels   0; Gaps   0;

Qy   61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
           | || |||| ||||| | || || |||| || || | | ||||| ||| | || |
Db   129 AACAGGAGCGCTGGGTTGTGGGTATTATGCTTTCAGTGACTTTGGTGAAATTCATGCTAA 188

Qy   121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
           || | || ||||| | ||| ||||||||| || | |||||| || ||||| ||||| |
Db   189 TGATTTATTGCCGCTCTTTTACTAATGAGATTGTAAAGGCCTATGCTCAGGATCATTTTT 248

Qy   181 tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
           | || || |||||| || | || || | | || || | || | |
Db   249 TTGATGTGATCACTAATGTGATTGGTCTAATTGCTGCACTTTTGGCCAATTATTTTGATG 308

Qy   241 ggtggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgc 300
           ||||||||| ||||||||| || || | | || ||||||||| || || ||| |
Db   309 ATTGGATGGATCCTGTTGGTGCTATCATTCTGGCTTTGTACACAATTCGCACATGGTCAA 368

Qy   301 gaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctga 360

```
              || ||| |||| || ||      |||| | || | || || | | ||||| ||
Db    369 TGACAGTGTTGGAAAATGTGAATTCACTTGTTGGAAGATCAGCTGCACCTGAGTATCTTC 428

Qy    361 cgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
              ||| || || ||| |   ||||||||||||| ||| |   |||||| ||||| || |
Db    429 AGAAACTTACATACCTCTGCTGGAACCACCACAAGGCTGTGAGGCACATTGACACAGTTC 488

Qy    421 gagcct 426
              |||| |
Db    489 GAGCTT 494
```

RESULT   9
BI422631/c
LOCUS   BI422631   674 bp   mRNA   EST   16-AUG-2001
DEFINITION EST533297 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEC71M3 5' end, mRNA sequence.
ACCESSION   BI422631
VERSION   BI422631.1 GI:15197206
KEYWORDS   EST.
SOURCE   tomato.
  ORGANISM Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1 (bases 1 to 674)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE   Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT   Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES   Location/Qualifiers
     source   1. .674
                    /organism="Lycopersicon esculentum"
                    /cultivar="TA496"
                    /db_xref="taxon:4081"
                    /clone="cLEC71M3"
                    /clone_lib="tomato callus, TAMU"
                    /tissue_type="callus"
                    /dev_stage="25-40 days old"
                    /lab_host="XL1-Blue MRF'"
                    /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                    Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
                    of seedlings 7-10 days post-germination were excised, cut
                    at both ends and placed on MS medium with no selection.
                    Mixed callus was harvested at 25 and 40 days and included
                    undifferentiated masses. Tomato Callus EST Library"
BASE COUNT   207 a   131 c   140 g   196 t
ORIGIN

```
Query Match                 33.2%;   Score 141.4;  DB 11;   Length 674;
Best Local Similarity       66.7%;   Pred. No. 1.9e-28;
Matches 202;  Conservative     0;  Mismatches 101;  Indels    0;  Gaps    0;


Qy     124 tctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttcttcg 183
           I II III I      I      IIIIIIIII II I  II II IIIIIIIII IIIIIIIIII
Db     659 TGTATTGCGGTCTTTCACCCAATGAGATTGTTAAAGCATATGCCCAGGATCATTTCTTCG 600


Qy     184 acgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaatggt 243
           I II IIIIIIIII I I II II III IIII IIII III I  I II  I I
Db     599 ATGTTATCACAAACGTTATTGGACTAGTCGCGGCATTGCTTGCTAACTACTTCAGTGGCT 540


Qy     244 ggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgcgaa 303
           IIII IIIIIIIIIII II II  I  IIIIIIIIII II II   II III I     I
Db     539 GGATAGACCCTGTTGGAGCTATGATTCTCGCGTTGTATACCATTCGAACATGGTCAATGA 480


Qy     304 cggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctgacga 363
           I III I IIIIIIIII  I I II I II III I II  IIII II III  II
Db     479 CCGTGTTAGAGAACGTGAACTCTCTTGTCGGTAAGGCAGCTGCACCAGAATATCTGCAGA 420


Qy     364 agctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgcgag 423
           IIII II III I  IIIIIIIII II I I  II  IIII II II II III III
Db     419 AGCTGACTTACCTCTGCTGGAACCATCACAAAGCCATAAAGCATATAGATACAGTGAGAG 360


Qy     424 cct 426
           III
Db     359 CCT 357



RESULT   10
AW396729
LOCUS       AW396729      619 bp      mRNA            EST        07-FEB-2000
DEFINITION  sg80a05.y1 Gm-c1026 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1026-9 5' similar to TR:O80632 O80632 F12L6.11 PROTEIN. ;, mRNA
            sequence.
ACCESSION   AW396729
VERSION     AW396729.1  GI:6915132
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
    TITLE   Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
```

FEATURES   Location/Qualifiers
     source   1. .619
                        /organism="Glycine max"
                        /db_xref="taxon:3847"
                        /clone="GENOME SYSTEMS CLONE ID: Gm-c1026-9"
                        /clone_lib="Gm-c1026"
                        /tissue_type="Senescing leaves, mature plants, greenhouse
                        grown."
                        /lab_host="DH10B"
                        /note="Vector: pT7T3-Pac (Pharmacia); Site_1: EcoRI;
                        Site_2: HindIII; This cDNA library was constructed from
                        mRNA isolated from senecsing leave tissue of mature
                        greenhouse grown plants. Complementary DNA was synthesized
                        from mRNA using a 3' anchored poly(dT) primer. EcoRI
                        adapters were ligated to the blunt-ended cDNA fragments
                        followed by digestion with EcoRI and HindIII. The cDNA
                        fragments were directionally cloned into the EcoRI-HindIII
                        restriction site of the pT7T3-Pac vector. The ligated cDNA
                        fragments were transformed into DH10B host cells (Gibco
                        BRL). This library was constructed R. Shoemaker and J.
                        Erpelding."
BASE COUNT   154 a   125 c   140 g   200 t
ORIGIN

     Query Match   33.0%;   Score 140.4; DB 10; Length 619;
     Best Local Similarity   61.5%;   Pred. No. 3.5e-28;
     Matches 225; Conservative   0; Mismatches 141; Indels   0; Gaps   0;

Qy   61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
            | ||||||| |||| | || | | ||| || || | | ||||| ||| | || |
Db   93 AACAAGAGCGCTGGGTTGTGAGCATTATGCTTTCAGTGACTTTGGTGAAATTCCTGCTGA 152

Qy   121 tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct 180
            || | || || || | ||| | ||||||||| | | ||||| ||||||||| || || |
Db   153 TGATTTATTGTCGTTCTTTTACCAATGAGATTATTAAAGCCTATGCCCAGGATCACTTTT 212

Qy   181 tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat 240
            | || || ||||| || | ||||| | | || || | || |
Db   213 TTGATGTGATCACTAATGTCATTGGCCTTATTGCTGCACTTTTGGCAAATTATGTTGATG 272

Qy   241 ggtggatggaccctgttggcgccatactgatcgcgttgtacacgatcacgacgtgggcgc 300
            ||||||||||||||| || || || | | || ||||||||| || || ||| |
Db   273 ATTGGATGGACCCTGTCGGTGCTATCATTCTGGCTTTGTACACCATTCGCACATGGTCAA 332

Qy   301 gaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacctga 360
            || ||| |||| || || | ||| | || | || || | |||| || ||
Db   333 TGACAGTGTTGGAAAATGTTAATTCCCTGGTTGGAAGATCAGCAGCACCAGAATATCTTC 392

```
Qy    361 cgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacggtgc 420
          ||| || || ||| |   ||||||||||||| |||   |   |||||| || || || |
Db    393 AGAAACTTACATACCTATGCTGGAACCACCACAAGGCTGTGAGGCACATTGATACAGTTC 452


Qy    421 gagcct 426
          | || |
Db    453 GGGCAT 458



RESULT 11
AW756123
LOCUS       AW756123      507 bp    mRNA           EST        21-NOV-2000
DEFINITION  sl16b11.y1 Gm-c1036 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1036-1462 5' similar to TR:O80632 O80632 F12L6.11 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AW756123
VERSION     AW756123.1  GI:7685475
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 995   Std Error: 0.00
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1036-1462"
                     /clone_lib="Gm-c1036"
                     /tissue_type="somatic embryos cultured on MSD 20"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
```

cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restrictions site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT   132 a   104 c   114 g   157 t
ORIGIN

   Query Match   29.4%;   Score 125.2;   DB 10;   Length 507;
   Best Local Similarity 61.8%;   Pred. No. 4.7e-24;
   Matches 199; Conservative   0; Mismatches 123; Indels   0; Gaps   0;

Qy   105 gtgaagttcttcctcatgctctactgccgaacgttcaagaatgagatcgtgagggcctac 164
          ||||| ||| | || ||| | || || || | ||| | ||||||||| | | |||||
Db   1 GTGAAATTCCTGCTGATGATTTATTGTCGTTCTTTTACCAATGAGATTATTAAAGCCTAT 60

Qy   165 gcccaggaccatttcttcgacgtaatcacaaactctgtcggcctggtctcggcgctgctc 224
          ||||||||| || || || || || || |||||| || | ||||| | | || || |
Db   61 GCCCAGGATCACTTTTTTGATGTGATCACTAATGTCATTGGCCTTATTGCTGCACTTTTG 120

Qy   225 gctgtccggtacaaatggtggatggaccctgttggcgccatactgatcgcgttgtacacg 284
          || | |||||||||||||| || || || | | | || ||||||||
Db   121 GCAAATTATGTTGATGATTGGATGGACCCTGTCGGTGCTATCATTCTGGCTTTGTACACC 180

Qy   285 atcacgacgtgggcgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcg 344
          || || ||| | || ||| ||||||| || || | ||| | || | || ||
Db   181 ATTCGCACATGGTCAATGACAGTGTTGGAAAATGTTAATTCCCTGGTTGGAAGATCAGCA 240

Qy   345 ccggcagagtacctgacgaagctcacgtacttgatctggaaccaccatgaggagatccag 404
          | |||| || || ||| || || || ||| | |||||||||||||| ||| | |
Db   241 GCACCAGAATATCTTCAGAAACTTACATACCTATGCTGGAACCACCACAAGGCTGTGAGG 300

Qy   405 cacatcgacacggtgcgagcct 426
          ||||| || || || || || |
Db   301 CACATTGATACAGTTCGGGCAT 322

RESULT   12
BG154726
LOCUS   BG154726   597 bp   mRNA   EST   06-FEB-2001
DEFINITION sab38c04.y1 Gm-c1026 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1026-3943 5' similar to TR:O80632 O80632 F12L6.11 PROTEIN.
            ;, mRNA sequence.
ACCESSION   BG154726
VERSION   BG154726.1 GI:12688390
KEYWORDS   EST.
SOURCE   soybean.

```
ORGANISM   Glycine max
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
           Glycine.
REFERENCE  1  (bases 1 to 597)
  AUTHORS  Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
           ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
           Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
           ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
           ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
           ,R., Waterston,R. and Wilson,R.
    TITLE  Public Soybean EST Project
  JOURNAL  Unpublished (1999)
  COMMENT  Contact: Shoemaker R/Public Soybean EST Project
           Public Soybean EST Project
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available through: Genome Systems, Inc. 4633 World
           Parkway Circle St. Louis, Missouri 63134 For further information
           call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
           427-3324 or contact: clones@genomesystems.com or
           info@genomesystems.com web site: www.genomesystems.com
           High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1026-3943"
                     /clone_lib="Gm-c1026"
                     /tissue_type="Senescing leaves, mature plants, greenhouse
                     grown."
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac (Pharmacia); Site_1: EcoRI;
                     Site_2: HindIII; This cDNA library was constructed from
                     mRNA isolated from senecsing leave tissue of mature
                     greenhouse grown plants. Complementary DNA was synthesized
                     from mRNA using a 3' anchored poly(dT) primer. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by  digestion with EcoRI and HindIII. The cDNA
                     fragments were directionally cloned into the EcoRI-HindIII
                     restriction site of the pT7T3-Pac vector. The ligated cDNA
                     fragments were transformed into DH10B host cells  (Gibco
                     BRL). This library was constructed R. Shoemaker and J.
                     Erpelding."
BASE COUNT      152 a     120 c     137 g     187 t      1 others
ORIGIN
```

```
  Query Match              28.5%;  Score 121.6;  DB 11;  Length 597;
  Best Local Similarity    59.9%;  Pred. No. 4.8e-23;
  Matches 221; Conservative    0; Mismatches 145; Indels    3; Gaps    1;

Qy     61 agcaagagctctggatggtcattagcatgtcctctgttgcggtcgtgaagttcttcctca 120
```

```
           |  |||||||| |||| | || |   |   |||    || || |   | ||||| ||| | || |
Db     44  AACAAGAGCGCTGGGTTGTGAGCATTATGCTTTCAGTGACTTTGGTGAAATTCCTGCTGA  103

Qy    121  tgctctactgccgaacgttcaagaatgagatcgtgagggcctacgcccaggaccatttct  180
           || | || || || | || | |||||||| || | ||||| |||||||| || || |
Db    104  TGATTTATTGTCGTTCTTTTACCAATGAGATTATTAAAGCCTATGCCCAGGATCACTTTT  163

Qy    181  tcgacgtaatcacaaactctgtcggcctggtctcggcgctgctcgctgtccggtacaaat  240
           | || || ||||| ||    | ||||| |  | | || ||  | ||            |
Db    164  TTGATGTGATCACTAATGTCATTGGCCTTATTGCTGCACTTTTGGCAAATTATGTTGATG  223

Qy    241  ggtggatggaccctgttggcgccatactgatcgcgttgtacacgat---cacgacgtggg  297
           ||||||||||||||| || || ||  |  | || |||||||| ||    |||    | |
Db    224  ATTGGATGGACCCTGTCGGTGCTATCATTCTGGCTNTGTACACCATTCGCACATTGGGTA  283

Qy    298  cgcgaacggtgctggagaacgtaggcacactgataggcaagtcggcgccggcagagtacc  357
           || | | |   ||| || ||     | ||| | || |  || ||  |  |||| || |
Db    284  TTGGACAGTGGTGGGAAAATGTTAATTCCCTGGTTGGAAGATCAGCAGCACCAGAATATC  343

Qy    358  tgacgaagctcacgtacttgatctggaaccaccatgaggagatccagcacatcgacacgg  417
           |   ||| || || ||| |   |||||||||||| |||   |   | |||||| || || |
Db    344  TTCAGAAACTTACATACCTATGCTGGAACCACCACAAGGCTGTGAGGCACATTGATACAG  403

Qy    418  tgcgagcct  426
           | || || |
Db    404  TTCGGGCAT  412


RESULT  13
AU031216
LOCUS       AU031216       401 bp     mRNA            EST        29-OCT-1998
DEFINITION  AU031216 Rice cDNA from immature leaf including apical meristem
            Oryza sativa cDNA clone E61155_1A, mRNA sequence.
ACCESSION   AU031216
VERSION     AU031216.1  GI:3767106
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from immature leaf including apical meristem
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1..401
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
```

```
                    /db_xref="taxon:4530"
                    /clone="E61155_1A"
                    /clone_lib="Rice cDNA from immature leaf including apical
                    meristem"
                    /dev_stage="immature"
                    /note="Organ: leaf; immature leaf including apical
                    meristem (under long day condition)"
BASE COUNT       92 a       86 c      105 g      116 t         2 others
ORIGIN


    Query Match              27.6%;   Score 117.4;   DB 10;   Length 401;
    Best Local Similarity    58.9%;   Pred. No. 5.9e-22;
    Matches 202;  Conservative    0;   Mismatches 141;   Indels    0;   Gaps    0;


Qy     84 agcatgtcctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttcaag 143
          |  ||||    || |   | || ||||| |   ||||    | |||||| |||  |    |
Db      6 ATCATGCTTTCAGCAACTGTGGTGAAACTTGCCCTCTACATATACTGCAGAAGCTCAGGG 65 .


Qy    144 aatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactctgtc 203
          |||    || ||   ||| || ||  |||||||||| ||||||  ||  | || ||   |||
Db     66 AATAGCATTGTCCAGGCATATGCAAAGGACCATTACTTCGATGTCGTAACCAATGTTGTT 125


Qy    204 ggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggcgcc 263
          || | || | || |||| |   || |    ||||||||| |||||  || || || || ||
Db    126 GGTTTAGTGGCTGCTGTGCTTGGAGATAAGTTCTTCTGGTGGATTGACCCAGTAGGGGCT 185


Qy    264 atactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgtaggc 323
           | ||  | || |||| || ||   || |||  |||| |  ||| ||    || || | ||
Db    186 GTGCTACTTGCTGTGTATACCATTGTGAATTGGTCTGGAACTGTATACGAAAATGCAGTT 245


Qy    324 acactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatctgg 383
          ||||||  | ||  |||  || || ||||      |||  ||| || || ||| |    | |
Db    246 ACACTGGTGGGTCAGTGTGCCCCTTCAGATATGCTGCAGAAACTGACATACCTCGCCATG 305


Qy    384 aaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
          ||  ||| ||     |    ||   | | ||||||||| ||||| |
Db    306 AAGCACGATCCACGTGTGAGGCGGGTTGACACGGTTCGAGCTT 348



RESULT  14
BG887449
LOCUS        BG887449       374 bp    mRNA            EST        30-MAY-2001
DEFINITION   EST513300 cSTD Solanum tuberosum cDNA clone cSTD5B19 5' sequence,
             mRNA sequence.
ACCESSION    BG887449
VERSION      BG887449.1  GI:14264535
KEYWORDS     EST.
SOURCE       potato.
   ORGANISM  Solanum tuberosum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE    1  (bases 1 to 374)
   AUTHORS   van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemingo,A.,
             Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
     TITLE   Generations of ESTs from dormant potato tubers
```

FEATURES   Location/Qualifiers
     source   1. .374
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cSTD5B19"
                     /clone_lib="cSTD"
                     /tissue_type="dormant tuber"
                     /dev_stage="one month post-harvest"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; This library targets genes expressed in dormant
                     tubers. This library was made from sections of dormant
                     tuber, avoiding the buds and epidermis. Tubers were stored
                     for one month post-harvest at 4oC. The tuber was peeled,
                     well away from the surface. Then it was chopped into 1-2
                     mm cubes and immediately frozen in liquid nitrogen. This
                     library is noted as P4 in Tanksley lab notebooks."
BASE COUNT   100 a   78 c   83 g   113 t
ORIGIN

  Query Match   27.1%;   Score 115.6; DB 11;   Length 374;
  Best Local Similarity 60.4%;   Pred. No. 1.8e-21;
  Matches 209; Conservative   0; Mismatches 134; Indels   3; Gaps   1;

Qy   81 attagcatgtcctctgttgcggtcgtgaagttcttcctcatgctctactgccgaacgttc 140
          | || |||| | | | || || || | ||| ||| |||||| ||| |
Db   9 ACTATCATGATAACGGCCACAGTGGTAAAACTTGCCCTTTGGCTATACTGCAGAAGCTCA 68

Qy   141 aagaatgagatcgtgagggcctacgcccaggaccatttcttcgacgtaatcacaaactct 200
          || | || || | ||||| || |||| || |||| |||| || ||||| | || ||
Db   69 GGAAACAACATTGTCCGTGCCTATGCAAAGGATCATTATTTTGACGTGGTTACTAATGTA 128

Qy   201 gtcggcctggtctcggcgctgctcgctgtccggtacaaatggtggatggaccctgttggc 260
          ||||| ||||| | || | || | | | | | ||| |||| || ||||||||
Db   129 GTCGGACTGGTAGCAGCTATACTTGGCGACAAATTCTACTGGGGGATTGATCCTGTTGGT 188

Qy   261 gccatactgatcgcgttgtacacgatcacgacgtgggcgcgaacggtgctggagaacgta 320
          ||||| | | ||| | || || ||||| | ||| |||| | |||| || || || | |
Db   189 GCCATTATTCTTGCACTTTATACCATCACCAACTGGTCAGGAACTGTTTTAGAAAATGCA 248

Qy   321 ggcacactgataggcaagtcggcgccggcagagtacctgacgaagctcacgtacttgatc 380
          | ||||| | || |||| || ||| | || ||| || ||| | ||||| | |
Db   249 GTGTCACTGGTGGGACAGTCAGCCCCGCCTGAATACTTGCAAAAGTTAACGTATCTTGTT 308

Qy   381 tggaaccaccatgaggagatccagcacatcgacacggtgcgagcct 426
          | |||| | | | || ||| || || || || ||||||||
Db   309 ATAAGACACCCTCAAGTGA---AGCGTATTGATACAGTTCGAGCCT 351

RESULT 15
BG588773

```
LOCUS        BG588773      729 bp     mRNA              EST        12-APR-2001
DEFINITION   EST490582 MHRP- Medicago truncatula cDNA clone pMHRP-57O22, mRNA
             sequence.
ACCESSION    BG588773
VERSION      BG588773.1  GI:13606913
KEYWORDS     EST.
SOURCE       barrel medic.
  ORGANISM   Medicago truncatula
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
             Medicago.
REFERENCE    1  (bases 1 to 729)
  AUTHORS    Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
             and Fraser,C.M.
  TITLE      ESTs from phosphate-starved roots of Medicago truncatula, 2001
  JOURNAL    Unpublished (2001)
COMMENT      Contact: Harrison M.J.
             Plant Biology Division
             The Samuel Roberts Noble Foundation
             2510 Sam Noble Parkway, Ardmore, OK 73401
             Tel: 580-223-5810
             Fax: 580-221-7380
             Email: mjharrison@noble.org
             The Samuel Roberts Noble Foundation: N387524e TIGR sequence name:
             MTHBD95TK More information is available at: http://www.medicago.org
             Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Medicago truncatula"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pMHRP-57O22"
                     /clone_lib="MHRP-"
                     /tissue_type="roots"
                     /dev_stage="phosphate-starved"
                     /lab_host="XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; At the trifoliate stage, M. truncatula plants were
                     transplanted to phosphate-free sand and grown for a
                     further 30 days. During  this period, they were fertilized
                     twice weekly with 1/2 Hoaglands  solutions containing 20uM
                     potassium phosphate. cDNA was prepared from polyA+
                     enriched RNA. The cDNA was directionally ligated into the
                     Unizap XR vector from Stratagene and packaged using
                     Gigapack III Gold packaging extracts. Plasmids containing
                     cDNA inserts  were excised from the recombinant lambda-Zap
                     phage using Ex-assist  helper phage and propagated in
                     XLOLR cells."
BASE COUNT        212 a     142 c     158 g     217 t
ORIGIN
```

```
   Query Match               25.3%;  Score 107.6;  DB 11;  Length 729;
   Best Local Similarity     57.9%;  Pred. No. 3.5e-19;
```